CHILDREN'S JUDGMENTS OF PERSONALITY ON THE BASIS OF VOICE QUALITY

By
JUDITH ANN PHILLIS

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL OF
THE UNIVERSITY OF FLORIDA
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA
1969

University of Florida
1853

Copyright by
Judith Ann Phillis
1969

To my parents

Acknowledgments

The author wishes to express her sincere gratitude for the assistance and guidance so willingly provided by Dr. James C. Dixon, Dr. Norman N. Markel, Dr. Henry S. Pennypacker, Dr. Madelaine Carey Ramey, Dr. Audrey S. Schumacher, and Dr. Vernon D. Van De Riet, the members of her doctoral committee.

The author is particularly grateful to the following committee members:

Dr. Norman N. Markel provided extensive assistance and encouragement with regard to both theoretical and practical problems.

Dr. Madelaine Carey Ramey's advice concerning matters of statistical design was most valuable.

Dr. James C. Dixon was a patient mentor through four years of graduate study.

The author also gratefully acknowledges the assistance of the following:

Mrs. Ruth Duncan arranged for the use of the facilities of the P. K. Yonge Laboratory School, and the teachers and students were extremely cooperative.

Mrs. Willa Selfrige devoted many hours to the preparation of the experimental tapes.

The facilities of the University Computing Center

and Mrs. Mary Lynch, consultant, made the analysis of the results a much simpler task.

Miss Kathy Wright provided much appreciated assistance in the tabulation of the data.

Miss Pat Powers carefully proofread the manuscript and Mrs. Charlotte Gregory typed it professionally despite the pressures of time.

TABLE OF CONTENTS

	Page
List of Tables	vii
List of Figures	ix
Chapter I - Introduction	1
Chapter II - Method	17
Chapter III - Results	27
Chapter IV - Discussion	60
Chapter V - Summary	71
Appendices	74
Appendix A	75
Appendix B	79
List of References	82

LIST OF TABLES

Page

Table 1 - Frequencies and Major Factor Loadings of the Adjective Scales 13

Table 2 - Original and Direct Magnitude Estimation T-scores..... 19

Table 3 - Means and Standard Deviations of Percentile Placement on Ability Tests for the Twelve Subject Groups..... 21

Table 4 - Means and Standard Deviations of Subject Age in Months for the Twelve Subject Groups..... 23

Table 5 - Subjects Discarded from the Twelve Subject Groups..... 24

Table 6 - Analysis of Variance of Mean Evaluative Ratings..... 28

Table 7 - Mean Evaluative Ratings of the Twelve Voices by the Males and the Females and by the Three Age Groups..... 29

Table 8 - Mean Evaluative Ratings of the Six Voice Types by the Males and the Females and by the Three Age Groups..... 35

Table 9 - Mean Evaluative Ratings of the Three Voice Quality Profiles by the Males and the Females and by the Three Age Groups..... 39

Table 10 - Analysis of Variance of Mean Dynamism Ratings..... 42

Table 11 - Mean Dynamism Ratings of the Twelve Voices by the Males and the Females and by the Three Age Groups..... 43

Table 12 - Significance Levels for the Newman-Keuls Tests on the Dynamism Means of the Voices of the Same Voice Type..... 51

Table 13 - Mean Dynamism Ratings of the Six Voice Types by the Males and the Females and by the Three Age Groups..... 52

Page

Table 14 - Mean Dynamism Ratings of the Three Voice Quality Profiles by the Males and the Females and by the Three Age Groups..... 56

LIST OF FIGURES

		Page
Figure 1	- The Sex X Voice Interaction of the Evaluative Analysis.....	30
Figure 2	- The Age X Voice Interaction of the Evaluative Analysis.....	31
Figure 3	- Schematic Presentation of the Newman-Keuls Tests on the Ordered Voice Evaluative Means.....	33
Figure 4	- Plot of the Voice Type Means at the Two Levels of A of the Evaluative Analysis.....	36
Figure 5	- Plot of the Voice Type Means at the Three Levels of B of the Evaluative Analysis.....	37
Figure 6	- Plot of the Voice Quality Profile Means at the Two Levels of A of the Evaluative Analysis.....	40
Figure 7	- Plot of the Voice Quality Profile Means at the Three Levels of B of the Evaluative Analysis.....	41
Figure 8	- The Sex X Voice Interaction of the Dynamism Analysis.....	44
Figure 9	- The Age X Voice Interaction of the Dynamism Analysis.....	45
Figure 10	- Schematic Presentation of the Newman-Keuls Tests on the Ordered Age Group Dynamism Means.	47
Figure 11	- Schematic Presentation of the Newman-Keuls Tests on the Ordered Voice Type Dynamism Means.....	49
Figure 12	- Plot of the Voice Type Means at the Two Levels of A of the Dynamism Analysis.....	53
Figure 13	- Plot of the Voice Type Means at the Three Levels of B of the Dynamism Analysis.....	54
Figure 14	- Plot of the Voice Quality Profile Means at the Two Levels of A of the Dynamism Analysis....	57
Figure 15	- Plot of the Voice Quality Profile Means at the Three Levels of B of the Dynamism Analysis..	58

Chapter I

Introduction

Voice and Personality

The manner in which something is said, as well as actual content, has long been recognized as extremely important in determining the message communicated. The American cowboy showed his belief in the veracity of this statement with the phrase, "Smile when you say that, Partner." Vocal variations, as well as gestural and facial ones, can also be employed to modify communicated messages. In addition, to the astute listener, information which is not necessarily communicated intentionally is available from paralinguistic aspects of the voice. Thus, by his manner of speaking, a lecturer may well inform his audience of his present mood and general personality characteristics as well as the topic on which he is speaking.

Despite the general agreement that the voice expresses both transient and relatively permanent states of the individual, it was not until Sapir's article, "Speech as a Personality Trait" (1927) appeared that systematic research in the area began. Classic studies in the area are those of Pear (1931) and Allport and Cantril (Allport and Cantril, 1934; Cantril and Allport, 1935). Findings

were that age, sex and total personality s[illegible]tches could be reliably matched with voices by the judges. In general, however, in matching many aspects of personality with voices the inter-judge reliability was much higher than the validity. This finding has been confirmed by recent research (Kramer, 1964). Kramer (1964) argues that the consistently high reliability in such judgments is itself a type of validity, and that the lack of agreement with criterion personality measures should cast doubt on the validity of these measures rather than on the validity of the voice as an indicator of individual differences.

In addition to the need for validity in the criterion measures of personality employed, Kramer (1963) suggests that the personalities of judges is a variable which should be investigated in voice and personality studies. In view of the fact that "...certain basic personality differences are a function of biological sex,..." Markel and Roblin (1965) investigated the effects of this variable. Differences in personality judgments were obtained on the evaluative dimension of a semantic differential as a function of sex-of-judge. These differences occurred as a function of content variations. In the Markel and Roblin study all voice samples were provided by the same speaker. Employing this procedure, no sex differences were found with regard to the activity and potency dimensions.

Another critical factor in these studies is the specification of the voices employed. Markel (1969a) suggests the use of the terms "structural" and "functional" to describe

the methods employed in analyzing speech. He notes that the functional approach consists of various types of rating techniques in which personality judgments are made directly by judges. These studies have been reviewed by Kramer (1963). The structural methods are those which employ some technique to quantify the speech without attempting to make judgments about the speaker. Studies employing the latter methods have been reviewed by Mahl and Schulze (1964). Dittman and Wynne (1961), in an attempt to develop a structural methodology, employed the typical idiographic approach of the linguist and were unable to obtain reliable ratings of paralinguistic phenomena. Utilizing a nomothetic approach (comparing the individual to a group rather than focusing on individual variations through time) Markel (1965) obtained high inter-judge reliability in the coding of three aspects of paralinguistic or non-phonemic phenomena: pitch, loudness and tempo.

Children as Judges in Voice and Personality Studies

Kramer (1964) has noted that while there has been much theorizing about the importance of developmental variables in the area of speech and personality, few empirical studies have been conducted. The primary contention in the area has been that young children are more sensitive to nonverbal vocal cues than are adults. Dimitrovsky (1964) found sex and age differences in children's ability to judge emotions on the basis of voice. In general, however, older children and females were more accurate in their judgments. Gates (1927),

in an earlier study, found that more intelligent and older children correctly identified the intended emotions of speakers more frequently. Goodenough's (1957) findings that an interest in other people is an extremely critical sex-typed behavior forgirls while boys are encouraged to develop an interest in objects is important in this regard. Especially relevant in predicting differential accuracy in judging personality from the voice is the finding that girls, more than boys, are encouraged to attend to the expressive behavior of others and utilize the information obtained in the manipulation of people. Thus, particularly when children are employed as judges, the sex and age of judges are important variables in the area of voice and personality.

The Voice Quality Profile

Markel's development of the Voice Quality Profile (VQP) resulted from his attempts to analyze voice qualities structurally rather than functionally. The advantage of such an approach is that the structural aspects of voice quality can then be related to a variety of both transient and relatively permanent personality characteristics. As mentioned above, the voice qualities on which reliability of ratings has been established are pitch, loudness and tempo. In order to determine the VQP of a particular speaker, a group of judges rates one or several voice samples of the individual in terms of these three dimensions. The mean of all ratings obtained on a dimension is the score of that dimension. The scores for all individual voices rated by

this technique are then converted to T-scores, one distribution of T-scores for each dimension. Three profile groups are then established and each individual is assigned to one on the basis of the highest T-score obtained. The three VQP groups are Peak-Pitch (P-P), Peak-Loudness (P-L) and Peak-Tempo (P-T).

With this technique, the VQP's of psychiatric patients (Markel, 1969a) and college students (Costanzo, Markel, and Costanzo, 1969; Vargas, 1968) have been obtained. Raters have always been graduate or undergraduate university students. The VQP has been related to MMPI profiles by Markel (1969a) and to perceived emotion by Costanzo, Markel and Costanzo (1969). Vargas (1968) has related the VQP to the frequency of positive experiencing and behaving, the California Personality Inventory, the Tennessee Self-Concept Scale, FIRO-B, self ratings of early childhood experiences, Jourard's Self-Disclosure Questionnaire and judge-rated self-disclosure. Considering the pattern which emerges from all three studies, Vargas (1968) hypothesized a two dimensional trait system to describe the personality differences among the three VQP groups.

The mean MMPI profiles (Markel, 1969a) of the three VQP groups indicate that P-P voices are associated with personality descriptions which include such adjectives as "withdrawn" and "introverted." Psychiatric patients with the P-P VQP are likely to have a poor prognosis and be diagnosed as psychotic. The P-L VQP among psychiatric patients is

indicative of a good prognosis and a diagnosis of depression or psychoneurosis. Personality descriptions of such people would likely include such adjectives as "immature" and "demanding." The P-T VQP is associated with an MMPI profile within normal limits, and thus, on the basis of the MMPI study, a personality description of healthy or average is suggested. The Costanzo, Markel and Costanzo (1969) study indicates that among college students P-P voices are more often associated with grief, P-L with anger or contempt and P-T with indifference regardless of the emotion intended by the speaker. This appears to support Markel's findings (1969a), if one considers that depression is often viewed as a means of dealing with anger and a grieving person is characteristically withdrawn. Notice that these are the findings of the two studies for the P-L and P-P voices, respectively. The relationship between indifference and normalacy is less easily understood; however, this may merely indicate that none of the other emotions was clearly portrayed. Essentially Vargas' (1968) study is in agreement with the previous two. He found that speakers whose VQP was P-T were significantly higher than the P-P and P-L speakers in self-disclosure, the reported frequency of positive experiencing and behaving and the number of positive early childhood experiences reported. All of these variables are associated with healthy, adjusted people. The Vargas study also indicates that the P-L VQP is associated with speakers who feel inadequate and dissatisfied and who are maladjusted in interpersonal relationships according to various

scales of the Tennessee Self Concept Scale. Scores on FIRO-B indicate that the P-L VQP among college males is characteristic of people who wish to control others but who are unwilling to give in return. FIRO-B scores also indicate significant differences between the P-P speakers and the P-L and P-T speakers in that the latter two groups seek to gain control in interpersonal relationships.

Thus, it appears that a consistent pattern has emerged from studies investigating personality and the VQP. The following is a summary of the behavioral correlates of the three structurally defined VQP's:

(1) P-P - withdrawn and reserved but capable of maintaining mutually rewarding interpersonal relationships with other people

(2) P-L - angry, demanding and seeking to control others but not giving in interpersonal relationships

(3) P-T - outgoing and seeking to manipulate others but also willing to give to others and able to maintain mutually rewarding relationships with them.

Selection of the Rating Technique

The problem of choosing a rating technique is a particularly difficult one when children are to be the raters. They are notorious for the unreliability of their ratings as well as for their tendency to adopt favorite response styles regardless of the stimulus properties to be judged. It is, of course, possible to control for such difficulties by employing either a ranking or a paired-comparisons method. However, neither of these approaches lends itself well to situations such as the present one,

where the to-be-rated stimuli are transient and where it is impossible for the child to have more than one stimulus before him at a time. Although this is a less serious drawback for the paired-comparisons approach, where only two stimuli are considered at once, the time required to obtain paired-comparison data for twelve stimuli would be prohibitive in a situation such as the present one where ratings on two dimensions are required.

The Semantic Differential

The Semantic Differential (Osgood, Suci and Tannenbaum, 1957) is a rating technique which provides simultaneous ratings on several dimensions of stimuli. The instrument was developed and tested extensively by Osgood et.al. as a technique to be employed in the measurement of connotative meaning. Semantic differential studies involve obtaining, from a number of subjects, ratings of several concepts or stimuli on a series of bipolar adjective scales. The consistently relevant dimensions of meaning have been determined by means of factor analyses which have been computed on the intercorrelations of the adjective scales. In the typical semantic differential study, the intercorrelations are calculated across stimuli and subjects. Such studies have consistently found that most of the common variance is accounted for by three dimensions or factors: evaluative (50% - 75%), potency (25%-30%), and activity (20%-25%) (Osgood, Suci and Tannenbaum, 1957). Typical evaluative scales are good-bad and friendly-unfriendly;

strong-weak and hard-soft are typical potency scales, and the activity dimension consists of scales such as fast-slow and active-passive.

Previous experiments have employed the semantic differential to obtain personality ratings on the basis of voice from adult judges (Markel, Meisels, and Houck, 1964; Markel and Roblin, 1965). Markel, Meisels, and Houck (1964) found that on the basis of voice alone schizophrenics were judged more potent than nonschizophrenics. Speakers whom the judges perceived as schizophrenic were judged more active than those perceived as nonschizophrenic. The authors interpret the failure to obtain differences on the evaluative dimension of the semantic differenital as attributable to the constant content employed by all speakers and to the similarity of voice set. They agree with Trager (1958) in defining voice set as those aspects of the voice attributable to societal group, sex, age, state of health and location. Markel and Roblin (1965) obtained differences on the evaluative dimension, but not on the activity and potency dimensions when judges were instructed to rate the personality of one speaker in three different content conditions. The sex-of-judge was also found to have a significant effect. The authors concluded that the evaluative dimension reflects a general attitude toward the speaker which results from the relationship between his voice set and what he says. The significance of sex-of-judge was attributed to sex differences in responding to inconsistency between voice set and content.

Children as subjects

Studies which have investigated the dimensions of meaning in children of various ages support the contention that it is a fairly consistent variable (Small, 1958; Brandwin, 1965; Lilly, 1965; Barnard, 1966; Di Vesta, 1966). It has been noted in several studies, however, that the activity and potency dimensions tend to coalesce when children are employed as raters (Small, 1958; Brandwin, 1965; Barnard, 1966). Maltz (1963) concluded that the meaning of concepts to children becomes stable sometime before the fourth grade in school is reached. Donahoe (1961) investigated the stability of the three dimensions separately and concluded that ratings with respect to activity stablize before, or during, the first grade in school; evaluative ratings stabilize prior to the third and potency ratings between the third, and sixth years in school. However, some caution must be employed in interpreting these latter two studies, since both assummed that the factor loadings of the scales found with adult subjects would also be obtained with children used as subjects. The findings of Di Vesta and Dick (1966) with respect to the reliability of the semantic differential at different age levels do seem to indicate that a general increase in the stability of children's semantic differential ratings occurs between the third and fourth grades. They found "...a definite increase in reliability of ratings made on the semantic differential between the third and fourth grades." Di Vesta and Dick conclude that if twenty children are employed in the rating procedure, the reliability of the

concept scores obtained should be satisfactory. They also report that, "All correlations are sufficiently high to indicate the functional utility of separate factor scores, based on only two scales for research purposes."

As is typical in studies in which the semantic differential is used to obtain ratings, it was necessary to consider several variables in choosing the scales for the present study. The criteria employed in selecting scales were as follows:

(1) All scales included were selected on the basis of their relevance to making ratings of the personalities of male speakers.

(2) The scales employed were constructed of adjectives with which children could be reasonably expected to be familiar.

(3) The scales were selected in a manner which assured that they loaded consistently on the dimension which they were selected to represent.

With respect to the third criterion, it was first necessary to determine which dimensions of meaning were relevant to the present study. As was mentioned previously, in most semantic differential studies, the first factor or dimensions to emerge have been evaluative, potency, and activity. Later factors have been given such names as oriented activity, stability, receptivity, pointedness, and novelty. It has already been noted that when children are employed as subjects in semantic differential studies the activity and potency dimensions merge into one second dimension which has frequently been labeled dynamism. Osgood, Suci, and Tannenbaum (1957, pp. 73-74) observe that the activity and

potency factors also coalesce into a dynamism factor when judgments are limited to sociopolitical concepts, that is, people and policies. Since the present study required children to make judgments of people exclusively, it was concluded that evaluative and dynamism were the most critical dimensions.

Scales were therefore selected for inclusion in the present study if, in previous studies, they had been found to load consistently on either the evaluative or the dynamism factors and to have small loadings on all other factors. In order to assure that the children would be familiar with all adjectives so selected, the word frequencies established by Thorndike and Lorge (1944) were utilized. Thorndike and Lorge recommend that words as frequent as twenty per million be taught to fourth graders, thus children of this age can be expected to be familiar with words of this, or greater, frequency. Table 1 lists the six pairs of bipolar adjectives employed in the present study as well as the frequency per million words of the selected adjectives according to the Thorndike and Lorge Word List and the major factor loadings of the scales according to previous studies which employed children as subjects.

Statement of the Problem

The purpose of the present study was to test directly the hypotheses emerging form the work with the VQP. It appears that the VQP groups which have been defined by a structural methodology can also be defined in functional

TABLE 1

Frequencies and Major Factor Loadings of the Adjective Scales

	Scales	Thorndike & Lorge Frequency per Million[1]		Major Factor Loadings			
				Barnard	Brandwin	Lilly	Small
1	Good-Bad	100+	100+	Eval.	Eval.	Eval.	Eval.
2	Fair-Unfair	100+	(100+)	Eval.	Eval.	Not Used	Not Used
3	Important-Not Important	100+	---	Not Used	Not Used	Eval.	Eval.
4	Strong-Weak	100+	50+	No Major Loading	Dyn.	Activity & Potency	Dyn.
5	Fast-Slow	100+	100+	No Major Loading	Dyn.	Activity	Dyn.
6	Large-Small	100	100	Not Used	Dyn.	Activity & Potency	Dyn.

[1]Frequencies in parentheses refer to frequency of the root word.

terms. Stated simply, if judges were asked to rate the personalities of speakers in each of the three VQP groups on the basis of voice quality alone, the personality descriptions obtained would be expected to differentiate among the three VQP groups by means of two behavioral or personality dimensions. The first dimension might be thought of as reflecting "aggressiveness" and includes those adjectives in the first phrase of the descriptions of personality correlates of the three VQP's. The second dimension includes those adjectives in the second portion of the descriptions of personality correlates of the three VQP's and might be summarized as "friendliness."

It was also believed that investigating the effects of low scores, as well as peak scores on the three qualities, would contribute significantly to an understanding of the relationship between voice and personality. Therefore, this variable was included in the design of the present study. In order to obtain additional information on the effects of individual differences among the judges, which Kramer has suggested as a critical variable, males and females at three age levels were employed as raters in the present study.

It should be noted that the evaluative and dynamism factors of the semantic differential are quite similar to the two dimensions emerging form the studies of personality and VQP. For example, dimension 1 was characterized by adjectives such as "reserved" and "withdrawn" at one pole and "demanding," "controlling," and "outgoing" at the other pole.

These adjectives appear to represent the meanings attached to the two poles of the dynamism dimension. Dimension 2 of the VQP personality descriptions included the ability to give and receive pleasure in relationships with people at one pole and dissatisfaction with human relationships at the other pole. It seems that ratings of people at the poles of dimension 2 would be represented by the poles of the evaluative dimension. For example, friendly - unfriendly is a frequent evaluative scale and in the discussion of dimension 2 it was referred to as a friendliness dimension.

On the basis of previous studies which investigated the relationship between personality and the VQP, the following hypotheses were framed and tested in the present study:

(1) Speakers rated as HT and HP on VQP will be rated as more positively valued by judges rating personality than speakers rated as HL on VQP.

(2) Speakers rated as HT and HL on VQP will be rated as more dynamic by judges rating personality than speakers rated as HP on VQP.

(3) Speakers whose VQP's are similar will also receive similar ratings on both the evaluative and the dynamism dimensions.

The results of the Dimitrovsky (1964) and Gates (1927) studies indicate that the above hypotheses are more likely to be confirmed among the female and the older subjects. Despite the empirical findings, however, the contention that younger children discriminate among nonverbal vocal cues better than do older children continues to survive. One reason for this is the fact that the methods employed to obtain the children's ratings were not extensively

documented for the age groups involved. It was one of the purposes of this study to investigate the effects of sex and age in judging personality from the voice.

Chapter II

Method

The Voices

The stimulus voices were obtained from raw data previously collected by Markel (1969a). In this study, each of six raters judged the pitch, loudness and tempo of seventy-eight voice samples twice. The original voice samples were obtained from psychiatric patients who were instructed to read the following sentence, "Many of the reports received last week showed a large increase in earnings." On the basis of these ratings, Markel was then able to divide the initial seventy-eight subjects into three VQP groups: P-P, P-L, and P-T. Each voice was assigned to the profile group which represented the voice quality for which the highest T-score had been obtained. These three groups differed significantly from each other both in terms of VQP and MMPI profile.

A possible criticism of Markel's approach is that it fails to evaluate the effects of extreme ratings on the three qualities. That is, a voice may be defined as "peak" on a quality yet obtain a relatively low absolute rating on that particular quality. In addition, the possible correlates of low ratings on the voice qualities are not evaluated. These considerations, and the need to explore the effects of

voice qualities with a smaller number of stimulus voices, led to the decision to select the voices for this study in the following manner: two high and two low voices for each quality were included. In order to be classified as high on quality A, a voice was required to obtain a T-score at least 1.5 standard deviations above the mean (T-score = 65) on quality A and between plus and minus 1.0 standard deviations (T-score = 40 - 60) on qualities B and C. The same requirements were employed in establishing other high and low stimulus voices.

In selecting the twelve voices from the original seventy-eight, an attempt was also made to match the high and low voices for a particular quality with respect to ratings on the other two qualities. That is, the two High Pitch (HP) voices were matched with the two Low Pitch (LP) voices with respect to loudness and tempo. Since the limited sample from which the twelve voices were chosen made this procedure difficult, the voices were divided into two sets, with the voices which were more closely matched being included in Set I. Markel (1969b), in an unpublished study, also obtained pitch, loudness and tempo ratings using a direct magnitude estimation technique. The correlations between the original and D.M.E. T-scores are above .85 for all three voice qualities. Table 2 presents the pitch, loudness, and tempo T-scores of the two sets of six voices which were obtained in the two rating conditions.

TABLE 2

Original and Direct Magnitude Estimation T-scores

Voice Type	Set Number	Original T-scores			D.M.E. T-scores		
		Pitch	Loudness	Tempo	Pitch	Loudness	Tempo
HP	1	72.55	46.09	49.24	67.41	49.52	56.78
LP	1	32.97	50.69	47.47	38.04	42.06	50.83
HP	2	66.76	49.76	54.55	66.24	47.25	54.49
LP	2	32.97	42.42	48.35	39.51	40.77	46.71
HL	1	47.45	72.74	56.33	47.05	73.16	56.78
LL	1	39.73	35.06	50.13	40.78	41.10	55.41
HL	2	44.55	70.90	41.26	63.49	78.67	42.58
LL	2	48.42	33.22	40.38	39.51	39.80	39.92
HT	1	57.11	47.00	70.50	59.92	45.31	75.10
LT	1	59.03	41.49	35.06	54.88	41.10	37.01
HT	2	51.32	51.60	67.84	57.62	50.49	65.48
LT	2	52.28	53.44	34.18	53.70	56.32	36.63

Subjects

Subjects for this study were students from the University of Florida Laboratory School. Approximately 50% of the children enrolled in the school are children of professors at the University of Florida. The remainder of the students are drawn from the city of Gainesville and other communities in Alachua County. The professional groups are highly represented in the parents of the children which may account for the fact that, as a group, the children in the Laboratory School score higher on various ability tests than other children in the state.

Ability test scores were obtained from the school records in order to balance the experimental groups with respect to this variable. In addition, birthdates were obtained for all subjects. Six of the experimental groups made judgments on the evaluative scales and six of them made dynamism ratings. The elementary school groups, two groups of males (ELM) and two groups of females (ELF), were composed of fourth and fifth graders; eighth and ninth graders were sampled to obtain the junior high males (JRM) and the junior high females (JRF); and eleventh and twelfth grade males and females were sampled to obtain the senior high males (SRM) and females (SRF).

Ability test mean percentile scores and standard deviations are presented in Table 3. The percentiles were based on national norms of the California Test of Mental Maturity for the fourth, fifth and eight grades and of the School and College Abilities Test for the ninth, eleventh

TABLE 3

Means and Standard Deviations of
Percentile Placement on Ability Tests
for the Twelve Subject Groups

	Evaluative			Dynamism		
	Mean	N	SD	Mean	N	SD
ELM	73.55	20	15.29	71.38	16	21.87
ELF	69.75	20	21.22	77.29	17	11.83
JRM	72.53	17	18.11	75.60	20	20.77
JRF	73.26	19	27.02	71.15	20	24.40
SRM	71.06	18	20.32	69.10	20	26.94
SRF	66.26	19	26.48	69.73	15	18.46

and twelfth grades. Table 4, presents the means and standard deviations of the ages of the twelve subject groups. Tables 3 and 4 present data only on subjects actually employed in the analysis. Ability test scores were not available for all subjects and age data was not available for one subject in the SRF group. Thus, in tables 3 and 4 the N of several groups is less than twenty. Table 5 presents the data with regard to discarded subjects in each of the twelve subject groups. As can be seen from the table, subjects were discarded from each of the groups on the basis of ability test percentile scores in order to equate the groups with respect to this variable.

Procedure

The semantic differential was administered to all children in one of their regular classrooms during the school day. There were two classrooms for each grade, with approximately thirty students in each class. One section of each grade rated all twelve voices on the three evaluative scales; the remaining section rated the twelve voices on the three dynamism scales. This approach limited the number of ratings which any individual subject made to thirty-six and limited the possible effects of boredom on the ratings. In addition, through such a procedure, the independence of evaluative and dynamism ratings was assured.

The form of the semantic differential employed required the children to rate all twelve voices on one scale before beginning the ratings on another. It was felt that

TABLE 4

Means and Standard Deviations of Subject Age in Months for the Twelve Subject Groups

	Evaluative			Dynamism		
	Mean	N	SD	Mean	N	SD
ELM	125.10	20	7.49	124.95	20	7.24
ELF	124.50	20	5.43	124.25	20	6.43
JRM	172.80	20	6.46	174.60	20	7.13
JRF	174.75	20	8.62	173.90	20	7.89
SRM	208.56	20	8.10	209.75	20	7.07
SRF	207.50	20	6.53	210.05	19	7.47

TABLE 5

Subjects Discarded from the Twelve Subject Groups

	Total	Discarded Due to Incomplete or Patterned Marking	Randomly Discarded To Obtain Equal N	Discarded on Basis of Ability Test Percentile	Resulting N
ELM	28	1	4	3	20
ELF	28	0	5	3	20
JRM	25	1	1	3	20
JRF	28	0	5	3	20
SRM	29	3	3	3	20
SRF	27	2	2	3	20
	165	7	20	18	120
ELM	26	0	3	3	20
ELF	24	0	1	3	20
JRM	31	0	8	3	20
JRF	27	0	4	3	20
SRM	28	0	5	3	20
SRF	23	0	0	3	20
	159	0	21	18	120

this approach would minimize any "halo effects" in the ratings and would allow the children to form and maintain the desired set in making the ratings. All instructions and the twelve stimulus voices were recorded on a tape prior to the experimental sessions to approximate identical conditions in all groups. The instructions were modified from those employed by Markel and Hamp (1960).

Twelve tapes were recorded in all. Each began with a "dubbed" recording of the instructions and was followed by three different, random orders of the twelve stimulus voices. Approximately ten seconds elapsed between the voices in each set of twelve in order to allow time for the voice to be rated. A different random order of the voices was constructed for each of the thirty-six experimental presentations of the voices by reference to a table of random numbers.

Another aspect of the presentation in which ordering effects were important was the inter- and intra-scale ordering. Since the three evaluative scales and the three dynamism scales were each presented to six groups of children, it was possible to present each of the six possible orderings of the scales to one of the six groups. To counterbalance with respect to intra-scale ordering, three groups made the evaluative and dynamism ratings with the positive pole of the bipolar scales presented on the left. Three of the six groups making each type of rating rated the voices on bipolar scales on which the positive pole was printed on the right.

Data Analysis

Numerical values of 1-7 were assigned to the seven categories of the six scales. The dynamic and positively valued poles of the scales were assigned the numerical value of one and the opposite poles were assigned the numerical value of 7. The means of the ratings of the twelve voices on the three scales were then computed for each subject. This resulted in the twelve scores obtained for each subject. One exception to this procedure was due to the inadvertent ommission of the LP_2 voice among the voices rated on the large-small scale by the ninth grade. Because of this oversight, the LP_2 mean was computed from only two scales for those ninth graders making ratings on the dynamism scales.

The data were analyzed in two analyses of variance. Each analysis was conducted as a 2 x 3 x 12 repeated measurements design with repeated measurements over the third factor. This design is discussed by Winer (1962, 337-347). The two levels of factor A were sex-of-judge, the three levels of factor B were the three age levels (EL, JR, and SR) and the twelve levels of factor C were the twelve voices.

Chapter III

Results

Evaluative Results

Table 6 presents the results of the analysis of variance of the evaluative ratings. As can be seen from the table, F tests on the C main effect and the AC and BC interactions were significant, while tests on other main effects and interactions were not. These findings indicate that the subjects employed in the present study discriminated among the twelve voices on the evaluative dimension. Further, the pattern of these discriminations varied with both the sex and age of the rater.

Table 7 presents the mean ratings of the twelve voices by the three age groups and by the males and females. The AC means are plotted in Figure 1 and the BC means are plotted in Figure 2. As can be seen in both figures, the differencesbetween ratings of voices of the same type were large. Also, contrary to the findings expected, were the highly positive ratings of the HL voices.

This finding is directly contradictory of hypothesis 1. The initial prediction was that HT and PP speakers would be ratedas more positively valued than HL speakers. Because of this, the planned comparison of the HT, PP and HL voices was not completed. The null hypothesis,

TABLE 6

Analysis of Variance of Mean Evaluative Ratings

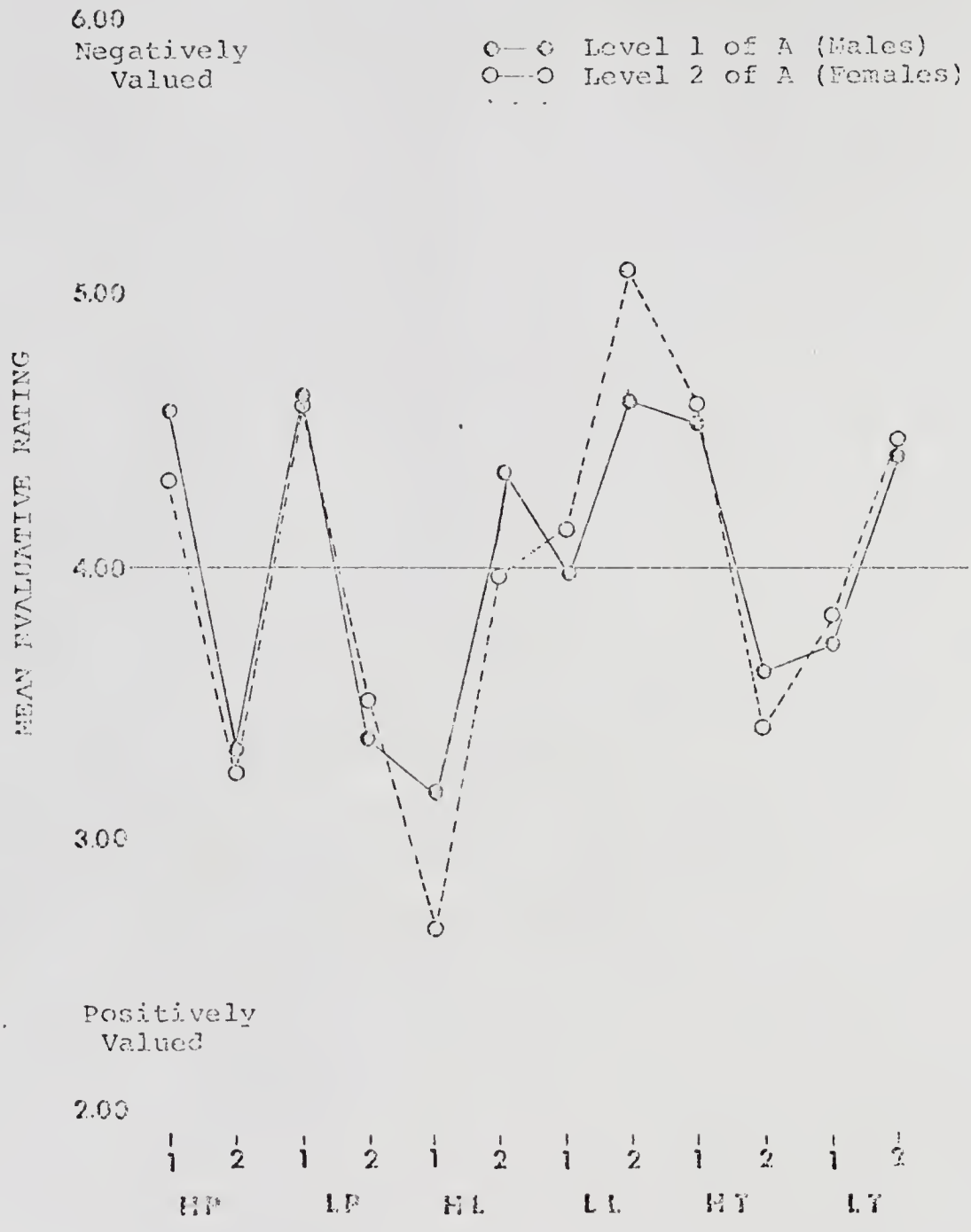
Source	Mean Square	df	F
Sex (A)	.429	1	.197
Age (B)	1.475	2	.680
AB	.376	2	.173
Error Between	2.170	114	
Voices (C)	44.096	11	46.281**
AC	1.761	11	1.848*
BC	3.885	22	4.078**
ABC	1.226	22	1.286
Error Within	.953	1254	

*significant at .05
**significant at .01

TABLE 7

Mean Evaluative Ratings of the Twelve Voices by the Males and the Females and by the Three Age Groups

	HP		LP		HL		LL		HT		LT	
	1	2	1	2	1	2	1	2	1	2	1	2
A_1 Males	4.53	3.32	4.62	3.39	3.17	4.36	3.96	4.61	4.51	3.61	3.71	4.40
A_2 Females	4.30	3.21	4.58	3.51	2.68	4.11	4.12	5.05	4.57	3.40	3.81	4.44
B_1 Elementary	4.25	2.98	4.69	3.89	2.88	3.70	4.32	5.03	4.53	3.59	4.28	4.48
B_2 Junior High	4.36	3.12	4.63	3.38	3.09	4.26	4.19	4.94	4.55	3.46	3.66	4.38
B_3 Senior High	4.63	3.69	4.47	3.09	2.91	4.74	3.62	4.52	4.53	3.46	3.34	4.39



The Twelve Voices

Figure 1

The Sex X Voice Interaction
of the Evaluative Analysis

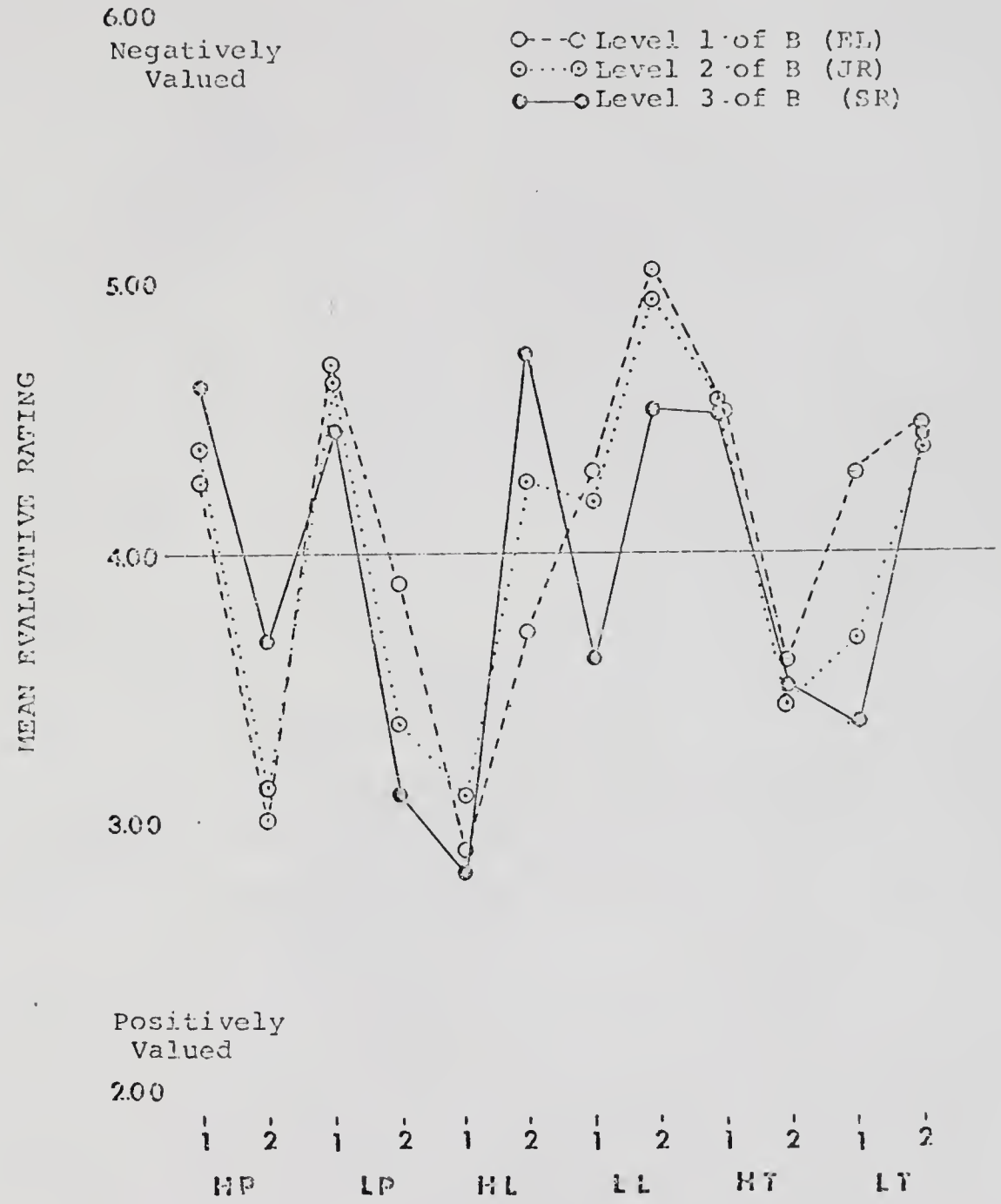


Figure 2

The Age X Voice Interaction
of the Evaluative Analysis

HO: $(\overline{HP}_1 + \overline{HP}_2 + \overline{HT}_1 + \overline{HT}_2) - 2(\overline{HL}_1 + \overline{HL}_2) \geq 0$, would obviously be retained at all levels of A and B.

Inspection of Figure 1 indicates that the sex by voice interaction was primarily a function of differences between the males and females in the ratings of the high and low loudness voices. Figure 2 indicates that the age by voice interaction also reflected differences in the ratings of the loudness voices and the HP and LT ratings as well. Figures 1 and 2 also show that the females and the elementary group tended to make the most extreme ratings of the HL and LL voices.

The Newman-Keuls technique (Winer, 1962; 309-310) was used to test differences among the ordered voice means at the two levels of A and three levels of B. These tests were conducted in order to evaluate hypothesis 3 which was that voices of similar voice type would also receive similar evaluative and dynamism ratings. The results of these tests are presented schematically in Figure 3. Voice means, which are underlined by common lines in this figure, were not found to be significantly different at the alpha level indicated. However, because differences among voices of the same voice type were obtained, it is difficult to interpret differences between voices of different voice types. As can be seen from Figure 3, of the thirty comparisons made between voices of the same type, all but three were significant at the .01 level. Thus, with respect to evaluative ratings hypothesis 3 must be rejected.

In order to evaluate the effect of voice type,

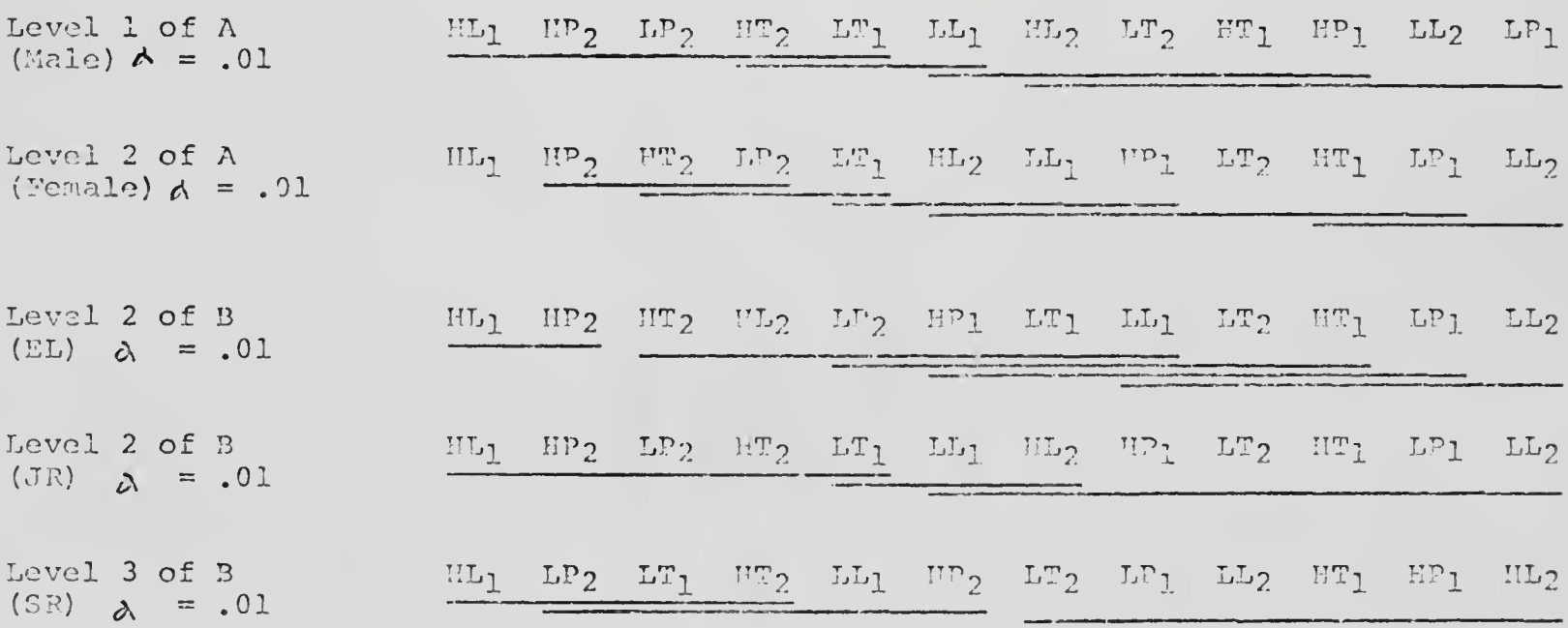

Level 1 of A
(Male) α = .01
HL1 HP2 LP2 HT2 LT1 LL1 HL2 LT2 HT1 HP1 LL2 LP1
Level 2 of A
(Female) α = .01
HL1 HP2 HT2 LP2 LT1 HL2 LL1 HP1 LT2 HT1 LP1 LL2
Level 2 of B
(EL) α = .01
HL1 HP2 HT2 HL2 LP2 HP1 LT1 LL1 LT2 HT1 LP1 LL2
Level 2 of B
(JR) α = .01
HL1 HP2 LP2 HT2 LT1 LL1 HL2 HP1 LT2 HT1 LP1 LL2
Level 3 of B
(SR) α = .01
HL1 LP2 LT1 HT2 LL1 HP2 LT2 LP1 LL2 HT1 HP1 HL2

Figure 3: Schematic Presentation of the Newman-Keuls Tests on the Ordered Voice Evaluative Means. (Means which share a common line were not significantly different at the α level indicated.)

despite the large differences between voices of the same type, the voice type means were calculated. Table 8 presents the voice type means for the males and females and for the three age groups. Figure 4 is a plot of the voice type means at the two levels of A; and the voice type means at the three levels of B are presented in Figure 5. Figure 4 shows that the evaluative ratings of both the males and the females discriminated most between the HL and the LL voice types. Both sexes rated HL voices toward the positive pole of the evaluative dimension and LL voices toward the negative pole. With regard to both the HL and the LL voice types, the ratings of the females were more extreme than those of the males. Differences between the high and low pitch and the high and low tempo voice types were much smaller, as were the differences between the male and female ratings of these four voice types.

As can be seen in Figure 5, the evaluative ratings of the six voice types display a clear developmental trend. In general, the EL age group discriminated most among the voice types, while the SR age group discriminated least. In all cases, the ratings of the JR age group fell between those of the EL and SR age groups. Figure 5 also shows that while the EL age group attached a negative evaluation to LP and LT speakers and a positive evaluation to HP speakers, just the reverse was true of the SR age group.

Because the evaluative analysis so clearly showed that the HT and HP voices were not more positively valued than the HL voices as originally hypothesized, an examination

TABLE 8

Mean Evaluative Ratings of the Six Voice Types by the Males and the Females and by the Three Age Groups

		The Six Voice Types					
		High Pitch	Low Pitch	High Loudness	Low Loudness	High Tempo	Low Tempo
A_1	Males	3.93	4.01	3.77	4.29	4.06	4.06
A_2	Females	3.76	4.05	3.40	4.59	3.99	4.13
B_1	Elementary	3.62	4.29	3.29	4.68	4.06	4.38
B_2	Junior High	3.74	4.01	3.68	4.57	4.01	4.02
B_3	Senior High	4.16	3.78	3.78	4.07	4.00	3.87

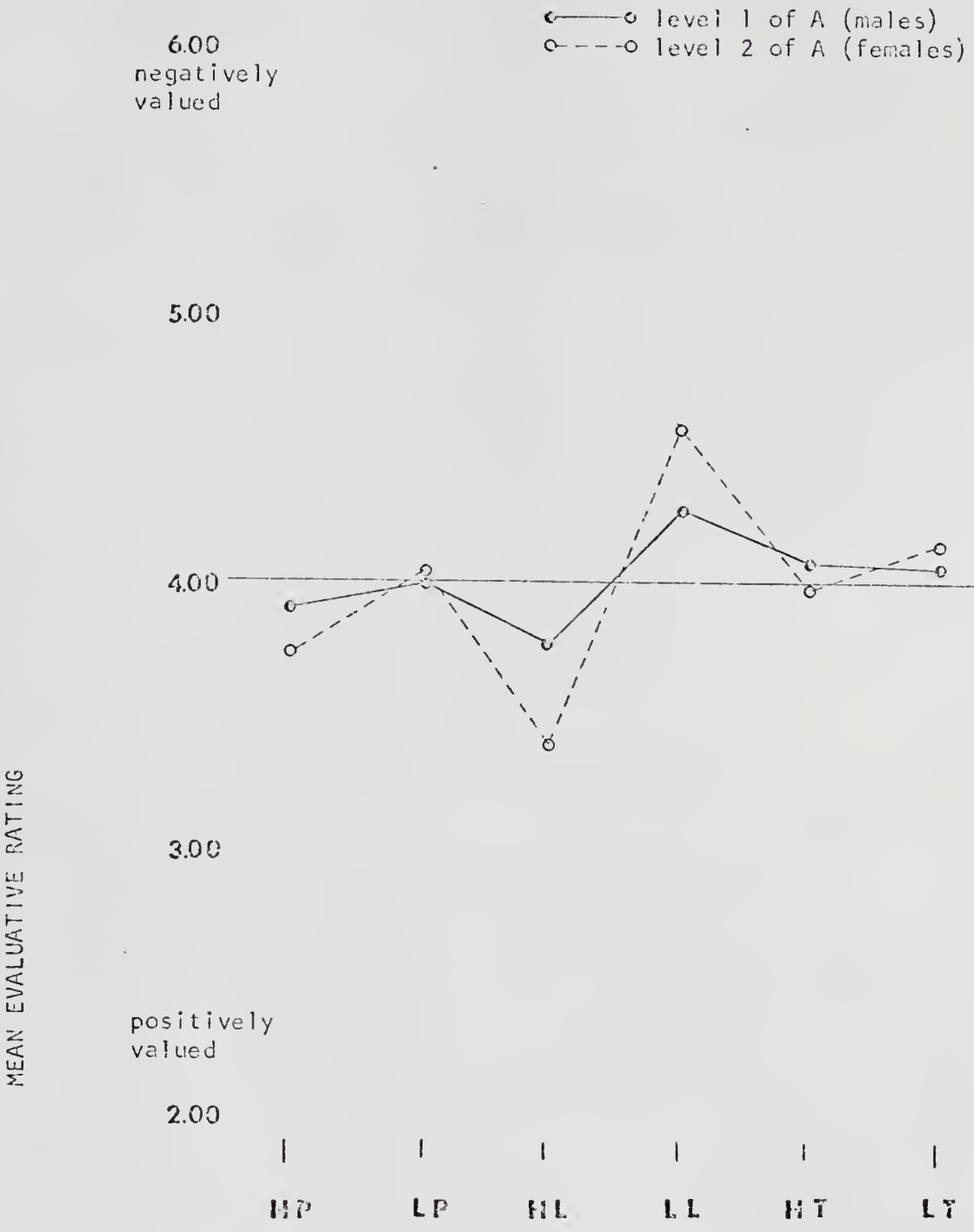


Figure 4

Plot of the Voice Type Means
at the Two Levels of A
of the Evaluative Analysis

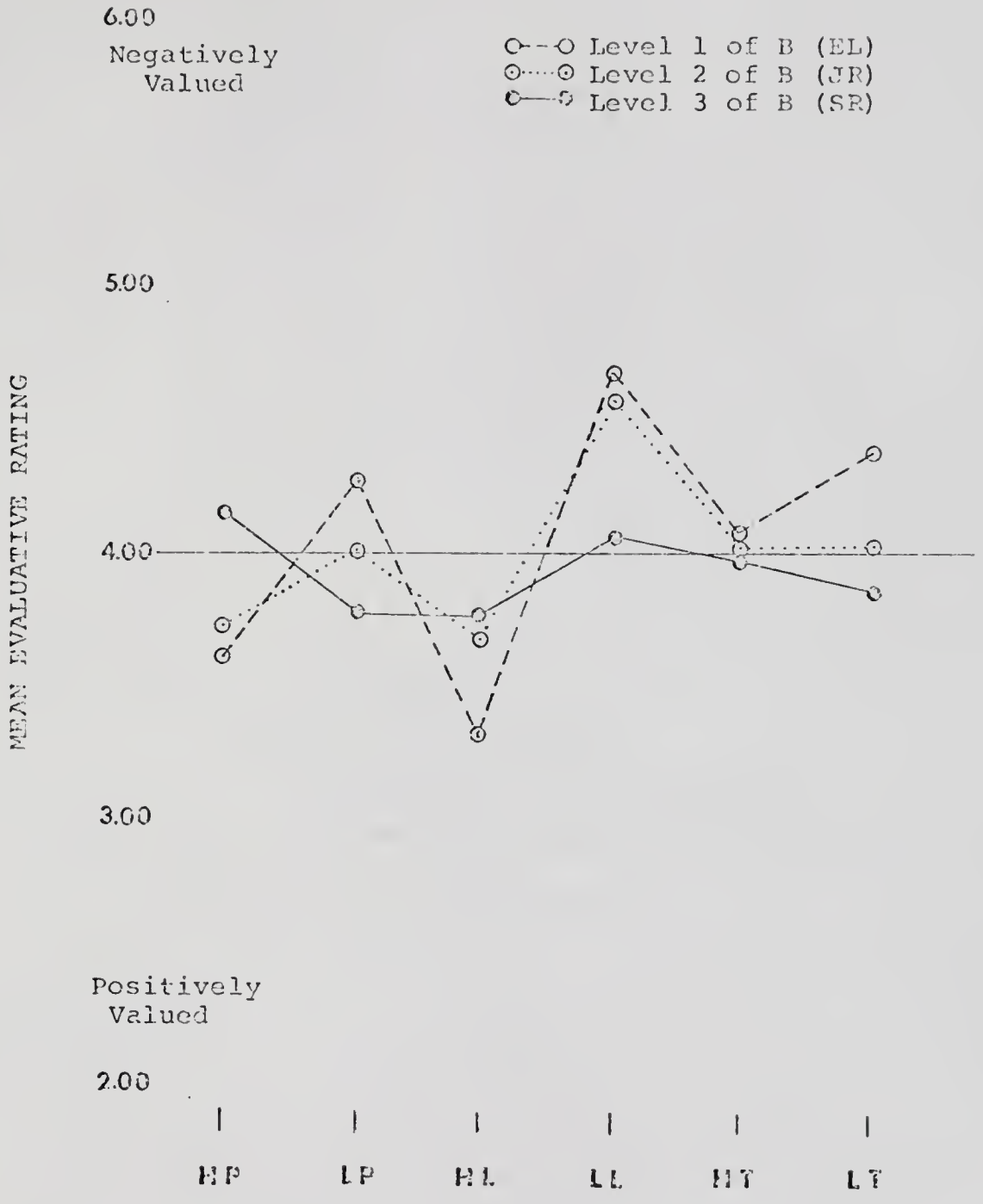


Figure 5

Plot of the Voice Type Means
at the Three Levels of B
of the Evaluative Analysis

of the voices in terms of voice quality profile (VQP) was made. Table 2 above shows that four voices of each VQP were included in the present study. The voices included in each profile group were as follows: P-P: HP_1, HP_2, LL_2 and LT_1; P-L: LP_1, HL_1, HL_2, LT_2; P-T: LP_2, LL_1, HT_1 and HT_2. Table 9 presents the mean evaluative ratings of the three VQP's made by the three age groups and by the males and females. Figure 6 is a plot of the VQP means at the two levels of A. The VQP means at the three levels of B are presented in Figure 7. Figures 6 and 7 show that, although differences among the VQP's were not large, among the males and the older children there was a tendency for the P-L profile voices to receive negative ratings. These same subject groups also tended to rate the P-T voices toward the positive pole of the evaluative dimension.

Dynamism Results

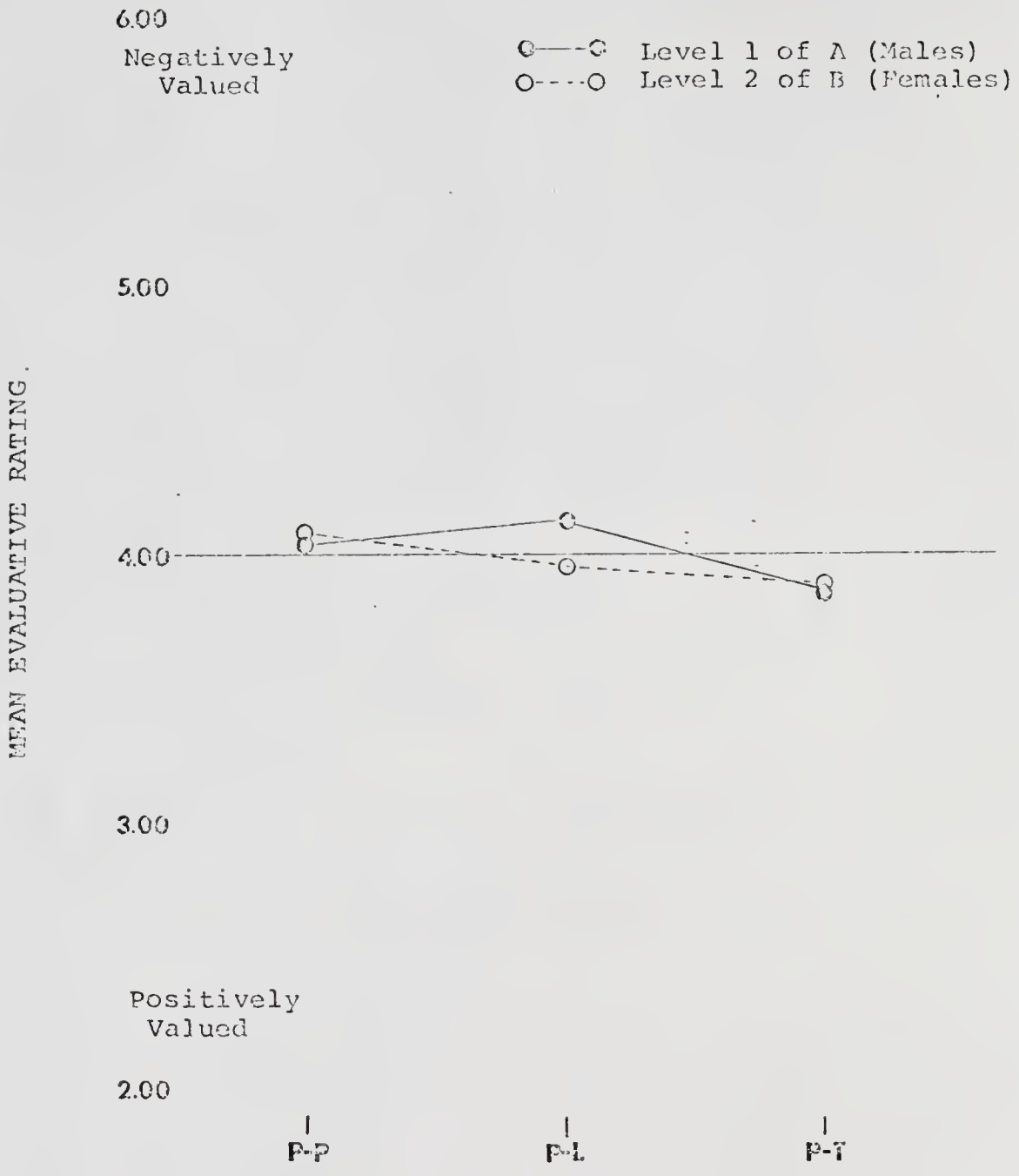
Table 10 presents the results of the analysis of variance of the dynamism ratings. As can be seen from the table, F tests on the B and C main effects and the AC and BC interactions were significant. These findings indicate that the pattern of discriminations among the voices on the basis of the dynamism dimension varied with both the age and sex of the rater.

Table 11 presents the means obtained from ratings of the twelve voices made by the three age groups and the males and females. The AC means are plotted in Figure 8 and the BC means are plotted in Figure 9. The figures show that

TABLE 9

Mean Evaluative Ratings of the Three Voice Quality Profiles by the Males and the Females and by the Three Age Groups

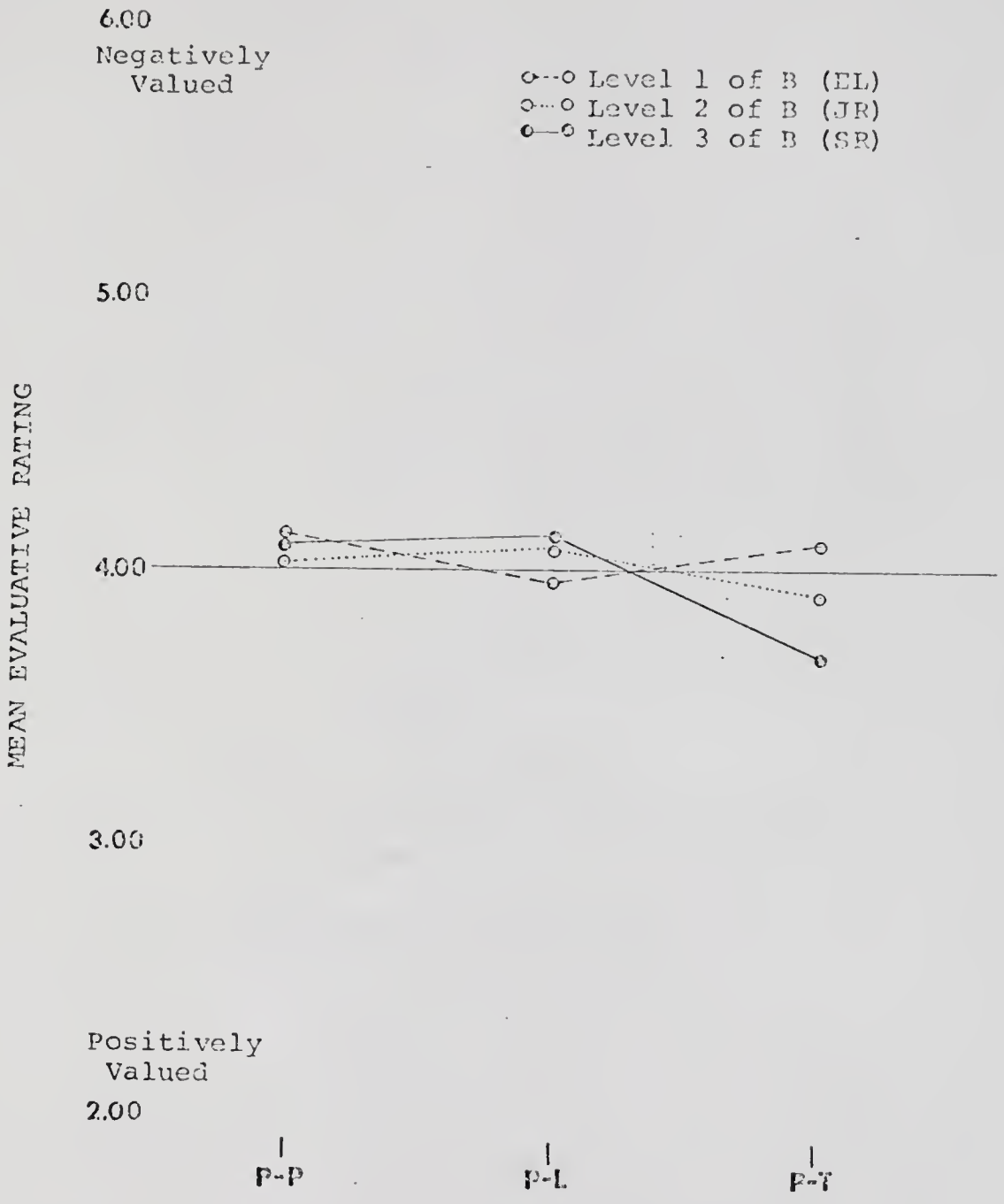
		The Three VQP Groups		
		Peak Pitch	Peak Loudness	Peak Tempo
A_1	Males	4.04	4.14	3.87
A_2	Females	4.09	3.95	3.90
B_1	Elementary	4.14	3.94	4.08
B_2	Junior High	4.02	4.09	3.90
B_3	Senior High	4.05	4.10	3.68



The Three Voice Quality Profiles

Figure 6

Plot of the Voice Quality Profile Means at the Two Levels of A of the Evaluative Analysis



The Three Voice Quality Profiles

Figure 7

Plot of the Voice Quality Profile Means at the Three Levels of B of the Evaluative Analysis

TABLE 10

Analysis of Variance of Mean Dynamism Ratings

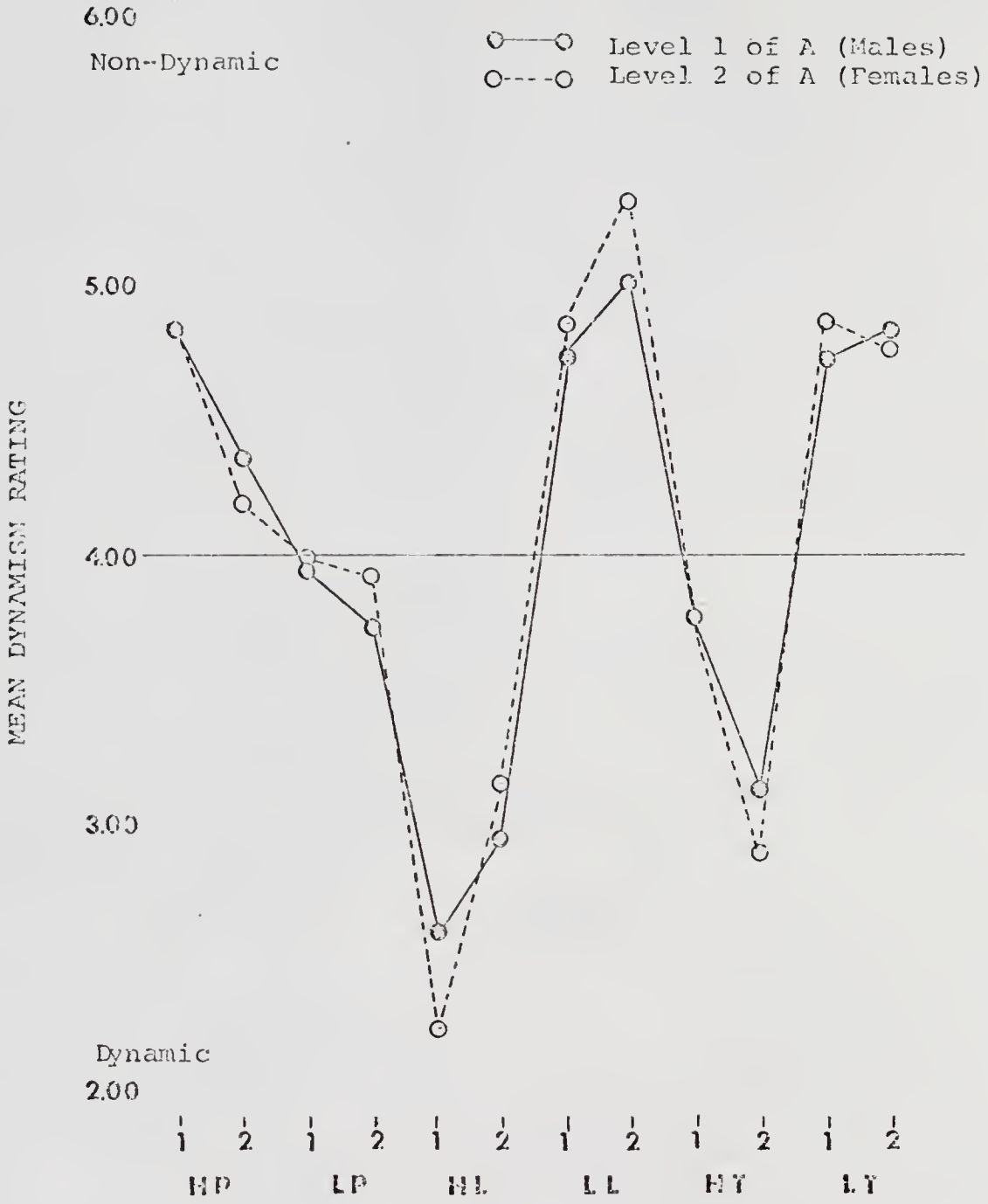
Source	Mean Square	df	F
Sex (A)	.299	1	.245
Age (B)	9.065	2	7.424**
AB	.675	2	.553
Error Between	1.221	114	
Voices (C)	91.518	11	168.232**
AC	1.060	11	1.949*
BC	3.306	22	6.079**
ABC	.809	22	1.487
Error Within	.544	1254	

*significant at .05
**significant at .01

TABLE 11

Mean Dynamism Ratings of the Twelve Voices by the Males and the Females and by the Three Age Groups

	The Twelve Voices											
	HP		LP		HL		LL		HT		LT	
	1	2	1	2	1	2	1	2	1	2	1	2
A_1 Males	4.85	4.35	3.97	3.77	2.62	2.95	4.75	5.03	3.82	3.12	4.74	4.84
A_2 Females	4.81	4.19	4.00	3.94	2.27	3.15	4.86	5.33	3.82	2.90	4.88	4.77
B_1 Elementary	4.77	4.11	3.35	3.54	2.24	2.51	5.14	5.23	3.82	2.88	4.69	4.73
B_2 Junior High	4.93	4.20	4.37	4.26	2.59	3.29	4.67	5.48	3.81	3.30	4.62	4.63
B_3 Senior High	4.78	4.50	4.24	3.76	2.51	3.35	4.62	4.82	3.83	2.87	5.13	5.06



The Twelve Voices

Figure 8

The Sex X Voice Interaction
of the Dynamism Analysis

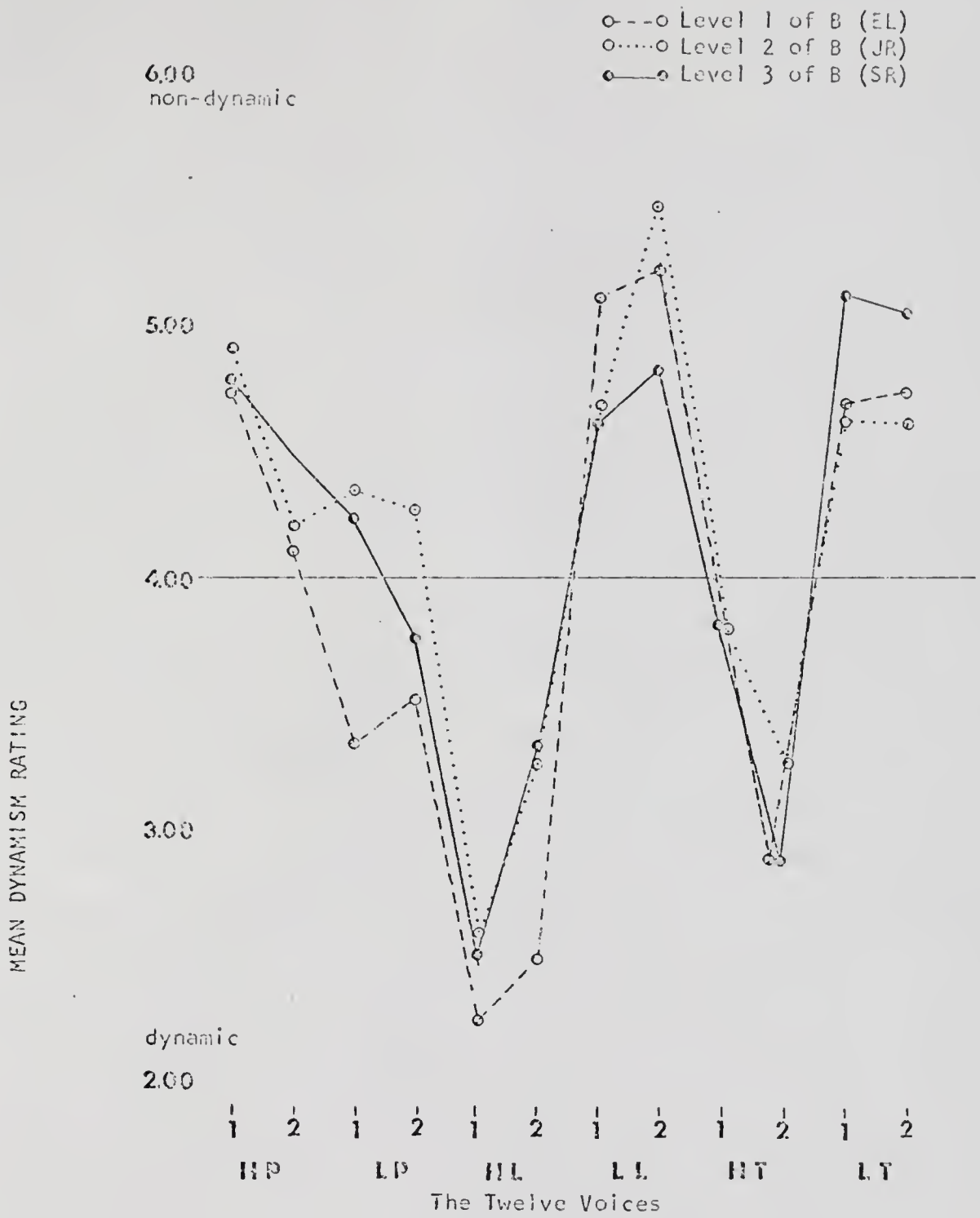


Figure 9

The Age X Voice Interaction of the Dynamism Analysis

intra-voice type differences were much smaller than inter-voice type differences. Inspection of Figure 8 indicates that the sex by voice type interaction was primarily attributable to differences between males and females in the rating of high and low loudness voices. There was an overall trend, however, for the females to make ratings, relative to the male ratings, toward the dynamic pole on the HP, HL and HT voices and toward the non-dynamic pole on the LP, LL and LT voices.

The plot of the age by voice interaction in Figure 9 shows that sizeable differences occurred among the age groups with regard to the ratings of several of the voices. The Newman-Keuls technique (Winer, 1962; 309-310) was used to test differences among the ordered age group means at each of the levels of C. The results of these tests are presented schematically in Figure 10. Age group means which share a common underlining were not found to be significantly different at the .05 level. As can be seen in Figure 10, significant differences among the age groups were found at eight of the twelve levels of C. Figure 9 shows that the EL age group tended to rate the HP, LP and HL voices more toward the dynamic pole than did either of the other two age groups. This tendency may have accounted for the significant B main effect. Figure 10 shows that, of the seven instances in which two of the age groups were significantly different from the remaining age group but not from each other, two involved differences between the JR age group and the other two. Although such results do

Level 1 of C B_1 B_3 B_2 (HP_1)

Level 2 of C B_1 B_2 B_3 (HP_2)

Level 3 of C B_1 B_3 B_2 (LP_1)

Level 4 of C B_1 B_3 B_2 (LP_2)

Level 5 of C B_1 B_3 B_2 (HL_1)

Level 6 of C B_1 B_2 B_3 (HL_2)

Level 7 of C B_3 B_2 B_1 (LL_1)

Level 8 of C B_3 B_1 B_2 (LL_2)

Level 9 of C B_2 B_1 B_3 (HT_1)

Level 10 of C B_3 B_1 B_2 (HT_2)

Level 11 of C B_2 B_1 B_3 (LT_1)

Level 12 of C B_2 B_1 B_3 (LT2)

Figure 10: Schematic Presentation of the Newman-Keuls Tests on The Ordered Age Group Dynamism Means

[Means Which Share a Common Line were not Significantly Different at the .05 alpha Level.]

not make a hypothesis of a developmental trend in the ratings untenable, neither do they offer strong support to such a hypothesis.

The significant AC and BC interactions necessitated evaluating hypothesis 2 by computing the planned comparisons among the dynamism ratings of the HP, HL and HT voices at the two levels of A and the three levels of B. Five t-tests were employed to test the following directional hypothesis: HO: $(\overline{HL}_1 + \overline{HL}_2 + \overline{HT}_1 + \overline{HT}_2) - 2\ (\overline{HP}_1 + \overline{HP}_2) \geq 0$. The null hypothesis was rejected at all levels of A and beyond the .001 level, one tailed. Thus, the high loudness and high tempo voices were rated by all groups as more dynamic than the high pitch voices.

The Newman-Keuls technique (Winer, 1962; 309-310) was used to test differences among the ordered voice type means at the two levels of A and the three levels of B. The results of these tests are presented schematically in Figure 11. Voice type means which share a common underlining were not found to be significantly different at the .01 level. Figure 11 shows that across all subject groups stable differences were found between high and low loudness and high and low tempo voices. In all cases, the HL and HT voices were rated as more dynamic that the LL and LT voices. The findings with regard to HP and LP voices were less consistent, but there was a general tendency for the LP voices to be rated as more dynamic than the HP voices.

As a means of evaluating hypothesis 3, which stated that voices of similar voice type would also receive similar

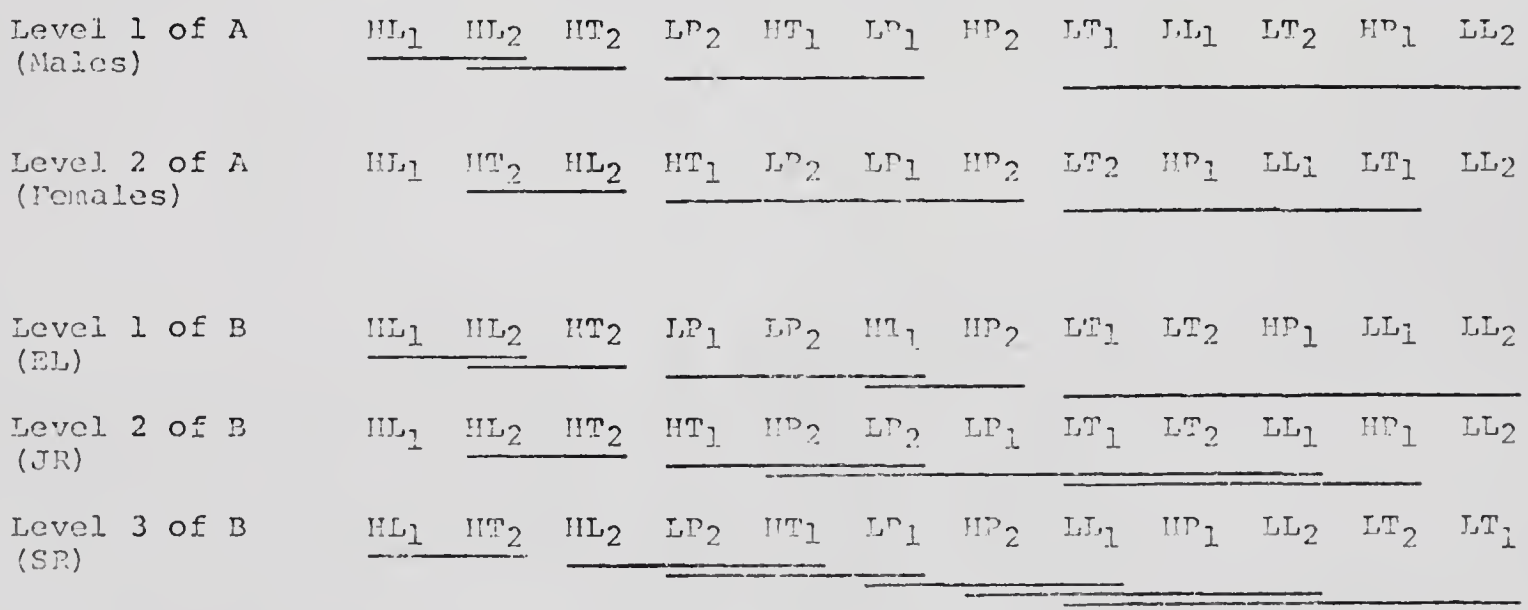

Level 1 of A (Males)
HL1 HL2 HT2 LP2 HT1 LP1 HP2 LT1 LL1 LT2 HP1 LL2
Level 2 of A (Females)
HL1 HT2 HL2 HT1 LP2 LP1 HP2 LT2 HP1 LL1 LT1 LL2
Level 1 of B (EL)
HL1 HL2 HT2 LP1 LP2 HT1 HP2 LT1 LT2 HP1 LL1 LL2
Level 2 of B (JR)
HL1 HL2 HT2 HT1 HP2 LP2 LP1 LT1 LT2 LL1 HP1 LL2
Level 3 of B (SR)
HL1 HT2 HL2 LP2 HT1 LP1 HP2 LL1 HP1 LL2 LT2 LT1

Figure 11: Schematic Presentation of the Newman-Keuls Tests on the Ordered Voice Type Dynamism Means. (Means which share a common line were not significantly different at the .01 level.)

evaluative and dynamism ratings, the results of the Newman-Keuls tests with respect to voices of the same type were examined. In order to state with more certainty that true differences between voices of the same type were not ignored, tests were also computed with alpha set at .10. Table 12 presents the significance levels obtained for the comparisons of voices of the same type at the two levels of A and the three levels of B. As can be seen from the Table, significant differences between the two HP, the two HL and the two HT voices were fairly consistently found across the levels of A and B. At least with regard to the dynamism ratings of these three voice types, the data of the present study did not support hypothesis 3. The data in Table 12, however, tended to confirm hypothesis 3 with regard to the LT and LP voice types and with regard to the EL age group.

In order to obtain a more reliable estimate of the contribution of voice type, the voice type means were calculated. Table 13 presents the voice type means for the males and females and for the three age groups. Figure 12 is a plot of the voice type means at the two levels of A; the voice type means at the three levels of B are plotted in Figure 13. Figure 12 shows that sex differences in the ratings were small and that both sexes rated HP, LL and LT voices toward the non-dynamic pole and LP, HL and HT voices toward the dynamic pole. For both sexes, the most extreme ratings were assigned to the high and low loudness voices,

TABLE 12

Significance Levels for the Newman-Keuls Tests on the Dynamism Means of the Voices of the Same Voice Type

	HP 1 and 2	LP 1 and 2	HL 1 and 2	LL 1 and 2	HT 1 and 2	LT 1 and 2
Level 1 of A (Males)	.01	NS	.10	NS	.01	NS
Level 2 of A (Females)	.01	NS	.01	.01	.01	NS
Level 1 of B (EL	.01	NS	NS	NS	.01	NS
Level 2 of B (JR)	.01	NS	.01	.01	.01	NS
Level 3 of B (SR)	NS	.10	.01	NS	.01	NS

TABLE 13

Mean Dynamism Ratings of the Six Voice Types by the Males and the Females and by the Three Age Groups

	The Six Voice Types					
	High Pitch	Low Pitch	High Loudness	Low Loudness	High Tempo	Low Tempo
A_1 Males	4.60	3.87	2.79	4.89	3.47	4.79
A_2 Females	4.50	3.97	2.71	5.10	3.36	4.83
B_1 Elementary	4.44	3.45	2.38	5.19	3.35	4.71
B_2 Junior High	4.57	4.32	2.94	5.08	3.56	4.63
B_3 Senior High	4.64	4.00	2.93	4.72	3.35	5.10

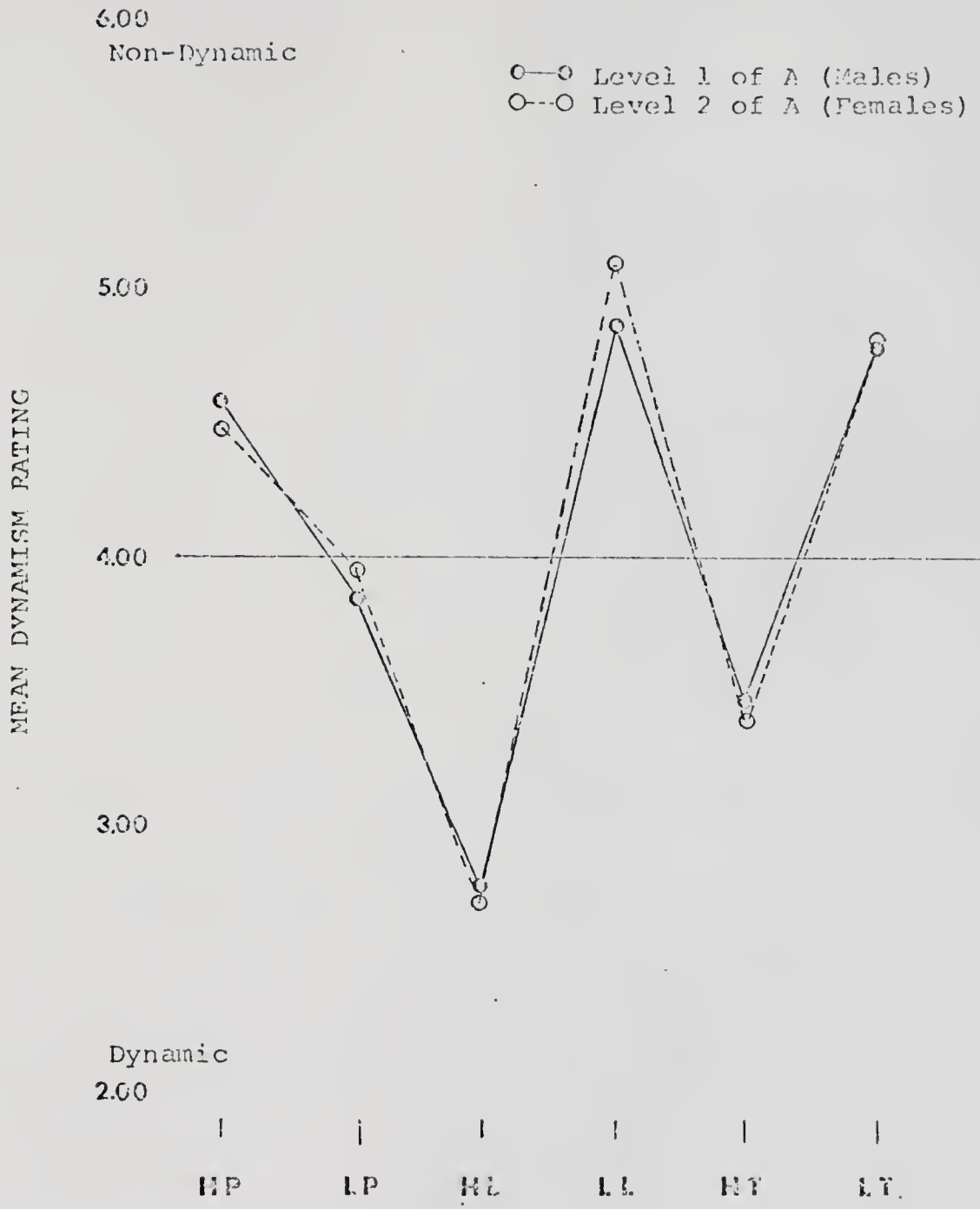
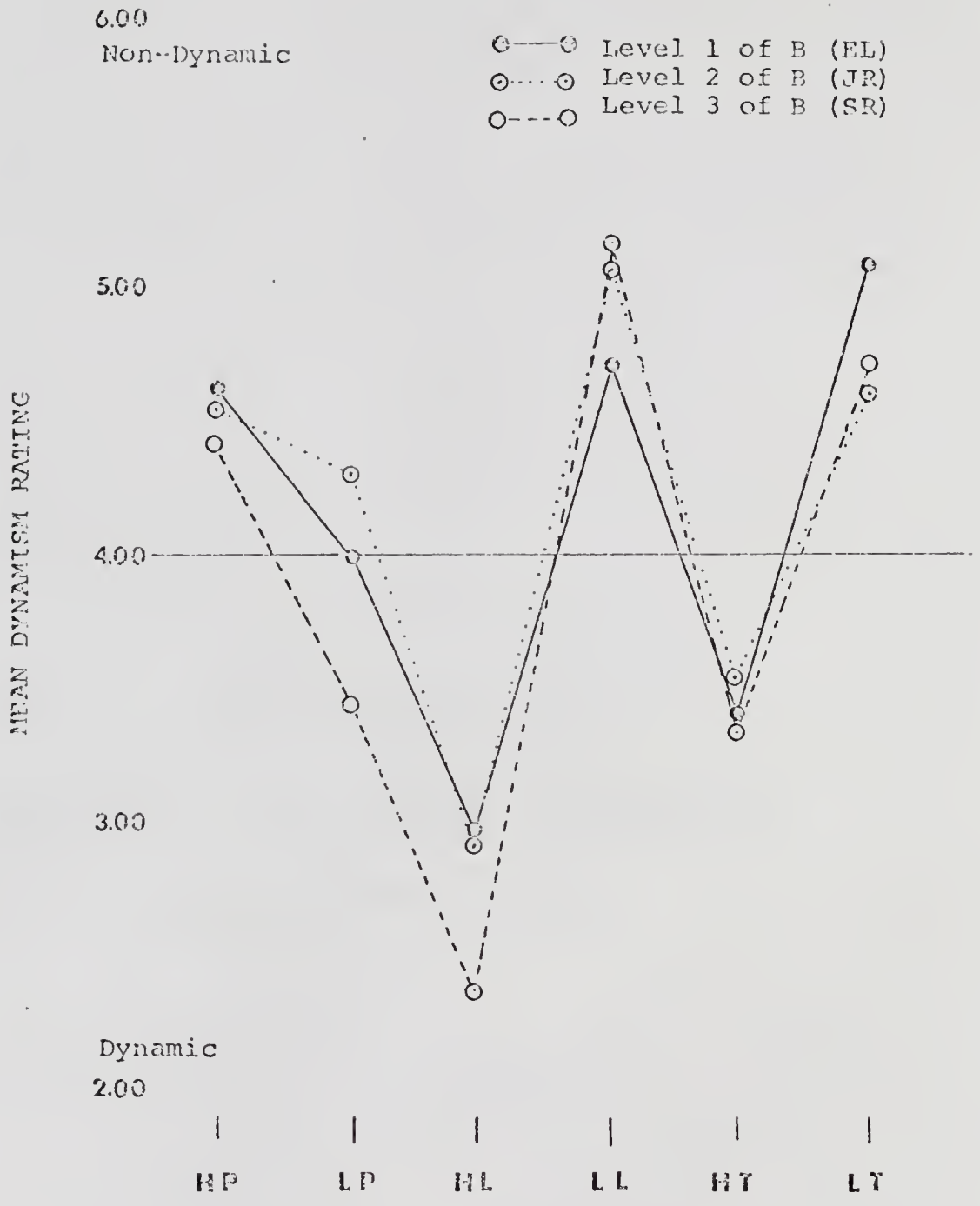


Figure 12

Plot of the Voice Type Means at the Two Levels of A of the Dynamism Analysis



The Six Voice Types

Figure 13

Plot of the Voice Type Means at the Three Levels of B of the Dynamism Analysis

while the least extreme ratings were assigned to the high and low pitch voices.

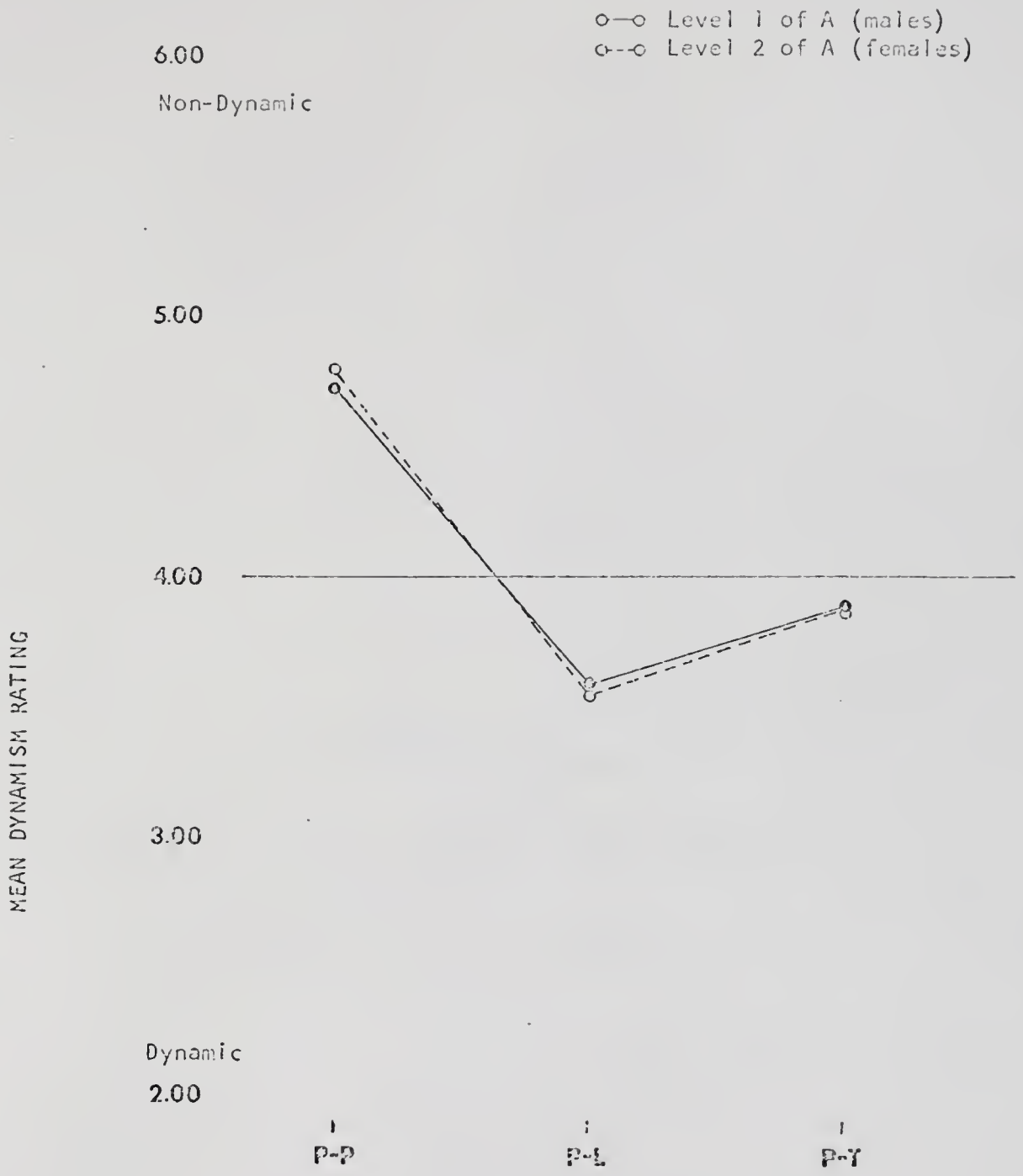
As can be seen in Figure 13, largest differences among the age groups occurred in the ratings of the LP, HL, LL and LT voice types. The LP voice type was rated toward the dynamic pole by the SP age group and toward the non-dynamic pole by the JR age group. Similarly to the males and females, the EL age group's rating of the LP voice type was near the dynamism scale midpoint. All groups rated the LL and LT voice types as non-dynamic and the HL voice type as dynamic. The ratings of these three voice types did show something of a developmental trend, in that differences occurred either between the two older and the youngest age groups or the two younger and the oldest age groups, but not between the middle age group and the other two.

It was also of some interest to consider the dynamism ratings in terms of Markel's original VQP procedure. Table 2 shows that four voices of each VQP were included in the present study. The voices included in each profile group were as follows: P-P: HP_1, HP_2, LL_2 and LT_1; P-L: LP_1, HL_1, HL_2, LT_2; P-T: LP_2, LL_1, HT_1 and HT_2. Table 14 presents the mean dynamism ratings of the three VQP's by the three age groups and the males and females. Figure 14 is a plot of the VQP means at the two levels of A; the VQP means at the three levels of B are presented in Figure 15. As can be seen in Figures 14 and 15, the P-P VQP was consistently rated as non-dynamic, the P-L VQP was consistently

TABLE 14

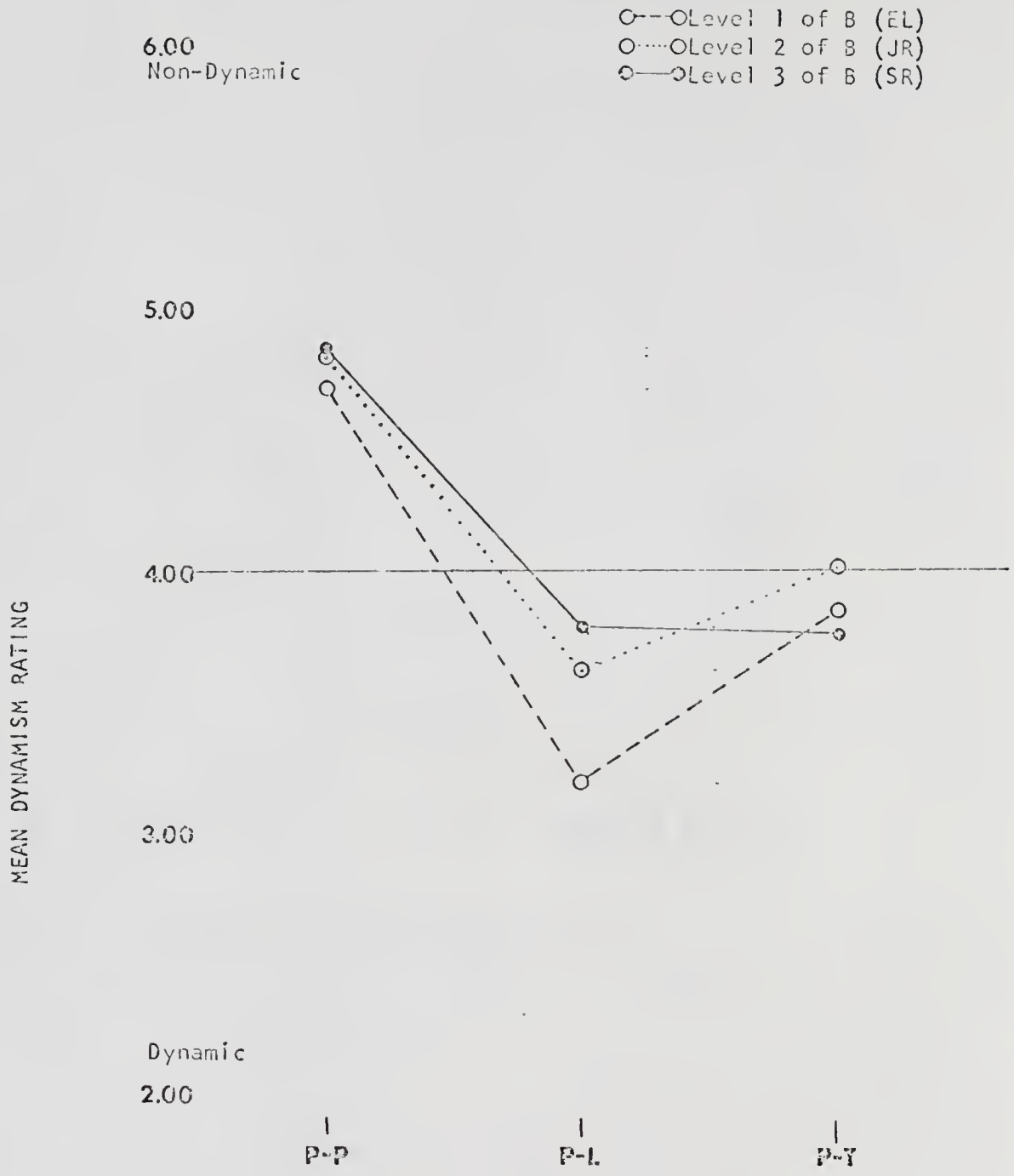
Mean Dynamism Ratings of the Three Voice Quality Profiles by the Males and the Females and by the Three Age Groups

		The Three VQP Groups		
		Peak Pitch	Peak Loudness	Peak Tempo
A_1	Males	4.74	3.60	3.87
A_2	Females	4.80	3.55	3.88
B_1	Elementary	4.70	3.21	3.85
B_2	Junior High	4.81	3.72	4.01
B_3	Senior High	4.81	3.79	3.77



The Three Voice Quality Profiles
Figure 14

Plot of the Voice Quality Profile Means
at the Two Levels of A of the
Dynamism Analysis



The Three Voice Quality Profiles

Figure 15

Plot of the Voice Quality Profile Means at the Three Levels of B of the Dynamism Analysis

rated as dynamic and the P-T VQP was generally rated as dynamic but less dynamic than the P-L VQP.

Chapter IV

Discussion

The extent to which the findings of any study can be generalized is dependent upon the subjects employed, the variables sampled and the methods used. In the present study, both males and females were employed as subjects and three different age levels of children were sampled. The population from which these children were drawn, however, included a high percentage of the children of professional people and, perhaps as a result, the children tended, as a group, to score high on standard ability tests. Stricker and Zax (1966) report that the ability to utilize all seven points of semantic differential scales is related to intelligence among children in the fourth and eight grades but not the twelfth grade. Thus, interpretations of findings with regard to the EL and JR groups are limited by this variable.

Other limiting factors in the present study include the technique employed to obtain the judgments and the limited sample of all male voices which the children rated. The findings of previous studies with regard to the evaluative and dynamism dimensions indicate that they are the most critical ones for obtaining judgments from children, especially when people are being rated. Previous studies

were also utilized to select individual scales. It should be noted however, that the results reflect only group stereotypes and do not provide information about individual subject responses to the separate speakers. In addition, it is not possible to make statements about all possible combinations of pitch, loudness and tempo nor about the possible effects of other voice qualities.

Sex Differences

The sex-of-judge affected both the dynamism and evaluative ratings. In both analyses, the main effect of sex was not significant, while the interaction of sex-of-judge with voice was. In general, this finding was attributable to the extreme ratings made by the females, especially of the HL and LL voices. Both males and females judged men who spoke loudly to be more dynamic and more positively valued than men who spoke softly.

Goodenough (1957) found that, in our culture, females, more than males, are encouraged to attend to the expressive behavior of others. That females would be more able than males to make discriminations of individual differences on the basis of voice alone, is a logical extrapolation from these findings. Dimitrovsky (1964) found this to be the case in a study which required children to select the intended emotion of speakers on the basis of voice alone. The correct-incorrect dichotomy employed by Dimitrovsky is not clearly applicable to the present study. However, the hypothesis, with regard to the dynamism ratings

of the six voice types was confirmed by the results of the present study. Thus, one way in which the ratings can be viewed as correct or incorrect is to consider the following as correct ratings: (1) dynamic: LP, HL and HT and (2) non-dynamic: HP, LL and LT. This system is essentially a restatement of the original hypothesis regarding the dynamism ratings. Reinspecting the voice type means reveals that, although all differences are small, the females' responses are extreme in the correct direction with regard to the HL, LL, HT and LT voice type and the responses of the males and extreme in the correct direction with regard to the HP and LP voice types. Thus, it appears that the females are more sensitive to loudness and tempo differences while the males are more sensitive to differences in pitch.

Sex differences in the evaluative ratings of the voices were much larger. Differences with regard to the pitch and tempo voice types were small, while the ratings of the two loudness voice types showed a definite interaction with the sex factor. The females made more extreme ratings than did the males of both the HL and the LL voice types.

Age Differences

The ratings of the voices on both the evaluative and the dynamism dimensions were affected by the age of the judge. In the evaluative analysis, the age by voice interaction was significant, while in the dynamism analysis both the interaction of age with voice and the main effect of age were significant. Further, the nature of the age

differences in the evaluative analysis was suggestive of a developmental trend in the ratings.

In practically all cases, the mean evaluative ratings of the voices by the middle age group (JR) separated the mean ratings of the voices made by the oldest (SR) and the youngest (EL) age groups. The few exceptions to this pattern occurred with respect to voices which elicited small differences in the ratings of the three age groups. When the data were converted to voice type means rather than individual voice means, the developmental pattern was clear across all voice types. All three age groups evaluated loud (HL) speakers as positively valued and quiet or soft (LL) speakers as negatively valued. None of the age groups judged the fast (HT) speakers to be either positively or negatively valued. However, the youngest children (EL) rated those who spoke slowly (LT) as negatively valued and the oldest children (SR) rated slow speakers (LT) as positively valued. Similarly, while the EL age group rated HP speakers as positively valued and LP speakers as negatively valued, just the reverse was true for the SR age group.

It seems possible that the difference obtained between the EL and SR age groups in the ratings of the HP and LP voices, might be attributable to the fact that elementary school children spend much of their time with females. Characteristically, the voices of females are higher in pitch than the voices of males and the EL age group rated HP voices as positively valued. Another explanation for the results is that even the boys in elementary school typically

have high pitched voices themselves. Perhaps the negative evaluation of the LT speakers by the EL age group reflects the characteristically higher activity level of children of this age. Older children, on the other hand, are typically more reflective and perhaps this accounts for the value they place upon a low tempo in speech.

Perhaps the most striking difference among the dynamism ratings of the three age groups was that the EL age group responded more consistently to voice type in judging the voices than did the other two age groups. The ratings of the former age group differentiated less between voices of the same voice type. This result seems somewhat inconsistent with the findings of both Gates (1927) and Dimitrovsky (1964). In both studies older children were superior to younger children in judging the intended emotion of the speaker on the basis of voice alone. As previously mentioned, a correct-incorrect dichotomy can be extrapolated from hypothesis 2 of the present study. The system requires that correct responses be defined as follows: (1) dynamic: LP, HL and HT and (2) non-dynamic: HP, LL and LT. Comparing the three age groups with the dichotomy in mind reveals that when significant differences occur among the age groups, the EL group's rating of the voices is always in the correct direction. Thus, in the present study, the youngest children employed were more often correct in making the dynamism ratings.

The Gates (1927) and Dimitrovsky (1964) studies, as did the present one, required the children to rate the

voices via a primarily verbal methodology. The methodologies employed in the earlier studies were developed specifically for that purpose. The semantic differential, the rating method employed in the present study, in contrast, has previously been extensively validated for the age groups included. Thus, it seems likely that younger children did not do as well as older children in the earlier studies because of an inability to express the discriminations which they made, rather than an inability to make the required discriminations. In fact, Gates (1927) noted that the youngest children in her study had difficulty understanding the instructions. In general, the present study offers empirical support to the theory that young children are particularly sensitive to nonverbal vocal cues.

Voice Type and Voice Quality Profile

One purpose of the present study was to determine if functional or personality ratings of voices would differentiate the voices in a pattern similar to the structural or voice quality ratings. The evaluative analysis revealed that although the subjects discriminated among the twelve voices, it could not be concluded that differences obtained resulted from differences among the six voice types. This was true because significant differences were consistently found between voices of the same voice type. The failure to obtain consistent differences on the evaluative dimension among voices of different voice types is in agreement with the findings of Markel and Roblin (1965). They found

differences on the evaluative dimension but not the activity and potency dimensions when voice quality and voice set were held constant. These differences, however, were attributable to content variations inconsistent with the speaker's voice set. Markel, Meisels, and Houck (1964) found no differences among speakers on the evaluative dimension when voice set and content were constant. Since content and voice set were constant in the present study, the differences obtained cannot be attributed to these variables.

In the Markel and Roblin study all ratings were made on just one speaker. Ratings in the Markel, Meisels and Houck study were mean ratings of groups. In the present study, on the other hand, the individual ratings of several speakers were employed. Thus, unlike the other studies discussed above, the present study could be affected by variations in voice qualities other than pitch, loudness and tempo.

An interesting result in the evaluative analysis was the clear and unexpected rating of the HL voices as positively valued and of the LL voices as negatively valued. It seems possible that this finding is related to the fact that all speakers were males. In our culture it is difficult for a man to be both dynamic and negatively valued. Also of interest was the finding that when the ratings were converted to three VQP means, the P-L VQP was rated as negatively valued by the males and the SR and JR age groups. This finding is in agreement with hypothesis 1 which

was not confirmed when the voice type data were examined.

With respect to the dynamism ratings the relationship between the structural ratings, and the functional ratings of the voices is more consistent with the original hypotheses. Although other variables perhaps other voice qualities, affected the dynamism ratings of the voices, few subject groups made intra-voice type differentiations among the LP and LT voice types. In addition, with the exception of the HP and HT voice tapes, the EL age group did not significantly differentiate between voices of the same voice type.

Hypothesis 2, that HL and HT voices would be rated as more dynamic than HP voices, was confirmed. Thus, not only did the dynamism ratings reflect a functional discrimination among the structural types but also the pattern of discriminations was that which was predicted from previous studies. That is, loud and fast speakers were rated as stronger, larger and faster than speakers whose voices were high in pitch. The most critical finding with regard to the LL and LT speakers was that they received ratings on the dynamism dimension which were just the reverse of the HL and HT ratings respectively. In addition, all subject groups rated the LP voice type as more dynamic than the HP voice type.

When the dynamism ratings were combined on the basis of the voice quality profiles, the pattern of ratings suggested by hypothesis 2 was clear. The P-L and P-T voices were rated as dynamic, while the P-P voices were rated as non-dynamic. With the exception of the SR age group, the

VQP means indicated that the P-L speakers were more dynamic than the P-T speakers. The SR age group differentiated very little between the P-L and P-T VQP groups.

In general, the subject groups consistently rated the HP and HL voice types good, fair and important, and the LL voice type bad, unfair and not important. The HL and HT voice types were rated strong, fast and large, and the HP, LL and LT voice types were rated weak, slow and small. Thus, males who speak loudly or with a high pitched voice are valued while males who speak softly are not, and males who speak loudly or quickly are seen as more dynamic than males who speak slowly, softly, or have high pitched voices.

Implications for Future Research

An interesting finding of the present study was the powerful effect of loudness on the ratings. On the basis of these findings, it seems likely that detailed investigations of loudness alone would be fruitful. In view of the fact that accurate, immediate feedback of intensity levels can be obtained mechanically, this variable could be investigated through many other methodologies.

The data also suggested that investigations of voice qualities in females would be a valuable contribution to the area. It was hypothesized that, in our culture, a dynamic man is also positively valued and that the evaluative ratings reflected this cultural bias. It seems logical to expect different patterns to emerge if female speakers are employed.

The ratings of the voices in the present study were affected by variables other than the voice qualities included. Since the study controlled for both voice set and content differences, it seems quite likely that the uncontrolled variables were other voice qualities. Especially with regard to evaluative ratings, the data of the present study suggest that these qualities may be extremely important.

There are a variety of reasons for investigating the relationship between voice and personality through methods other than the one employed in the present study. The semanticdifferential is not reliable for children younger than eight, yet this study indicates that young children are sensitive to differences in voice quality. Thus, a methodology which allowed the use of even younger subjects would be of value. It has previously been mentioned that the data of this study were group stereotypes of the different voice types. Although data regarding groups is an important aspect of psychology, a methodology which provided reliable information with regard to individual responses to the voice types would be desirable. Finally, of course, results which are independent of the methodology employed, are very unlikely to be spurious.

An approach which would deal simultaneously with several of the limitations of the methodology employed in the present study is to investigate voice and personality via an operant methodology. For example, loudness might be investigated by obtaining, from the same person, several speech samples identical in all respects except the loudness.

These stimuli could then be employed as subsequent events for as many operants on which a baseline had previously been obtained. Thus, in an operant study, changes in the rate of an operant would define the reinforcing properties of the different loudness levels. Such a methodology would make it possible to employ pre-schoolers, and even infants, in voice and personality studies.

Chapter V

Summary

One purpose of this study was to test directly hypotheses emerging from the work of Markel and his students on the voice quality profile (VQP). Six voice types were also defined in order to gain further information about high and low scores on pitch, loudness and tempo, the three qualities included in previous VQP studies. Thus, the six voice types included were as follows: high pitch (HP), low pitch (LP), high loudness (HL), low loudness (LL), high tempo (HT) and low tempo (LT). A second purpose of the study was to investigate age and sex differences in judging personality from the voice.

Two male speakers of each voice type were rated by all subjects. Three evaluative scales (e.g. "good-bad") and three dynamism scales (e.g. "large-small") from Osgood's Semantic Differential were used to obtain the ratings. The twelve voices were rated on the three evaluative scales by 120 males and females in the fourth, fifth, eighth, ninth, eleventh and twelfth grades. A similar group rated the twelve voices on the three dynamism scales.

The evaluative and dynamism ratings were analyzed in separate repeated measurements analyses of variance. Individual subject scores were the mean ratings made by the

subject ofeach of the twelve voices. The twelve voices were the levels of the repeated measurements factor; the levels of the two other factors were the three age groups and the two sex groups.

It was predicted that HP and HT speakers would be rated as more positively valued than HL speakers and that HL and HT speakers would be rated as more dynamic than HP speakers. The latter prediction was confirmed by the data while the former was not. Differences between voices of the same type led to the conclusion that voice qualities, other than pitch, loudness and tempo, affected the ratings, especially those made on the evaluative scales. In general, men who spoke loudly, or with high pitched voices, were rated as valued while those who spoke softly were not; men who spoke loudly or quickly were seen as more dynamic than speakers who spoke slowly or softly or with high-pitched voices. An interesting finding was that when the voices were examined in terms of VQF, rather than voice type, the data tended to confirm the hypothesis regarding the evaluative ratings as well as the hypothesis regarding the dynamism ratings.

The sex-of-judge affected both the evaluative and dynamism ratings of the voices. In the evaluative analysis, the sex differences were attributed to the extreme ratings of the females, especially of the HL and LL voices. The sex differencesin the dynamism ratings were attributed to the larger discrimination made by the males between the HP and LP voices. The females, on the other hand, discriminated

more between HL and LL voices and HT and LT voices than did the males.

Age differences were also found in both the evaluative and dynamism ratings. There was a consistent developmental trend in the ratings of the voices on the evaluative scales. With regard to the ratings of the voices on the dynamism scales, it was concluded that the youngest age group was the most accurate.

Further research was suggested with female speakers, other voice qualities, the loudness voice quality alone and younger judges.

APPENDICES

APPENDIX A

Instructions to Subjects

Everyone should now have a booklet containing four sheets. Raise your hand if you don't have a booklet or if your booklet doesn't have 4 sheets. (Pause) We are going to see if you can guess what some men are like just by hearing them say one sentence.

Before we try some examples please write the following information in the spaces provided on page 1 of your booklet. First your name (Pause); the month, day and year of your birth (Pause); your sex -- if you are a boy, put a circle around the M, if you are a girl, put a circle around the F (Pause); next fill in the grade you are in in school (Pause).

Now you are ready to learn how to make your guesses. You will always indicate what you think a man is like by making an X on an adjective scale. An adjective scale is two words which are opposite in meaning like tall and short. These words are printed on opposite sides of a page and are separated by 7 dashes or lines. Example A on page 1 is the adjective scale tall-short (Pause). Notice that the seven dashes or lines are separated by two dots. This is to help you place your X in the correct space. Example B shows the meaning of each of the seven spaces on the tall-short scale. You might like to read these labels

silently to yourself while I read them aloud. Putting an X in the space closest to the word tall would mean that you guessed the man was VERY TALL. The next space would indicate that the man was QUITE TALL. You would use this space if you thought a man was taller than SLIGHTLY TALL but shorter than VERY TALL. The space in the middle of the tall-short scale shouldbe used when your guess is that the man is AVERAGE in height. The first space on the short side of the tall--short scale should be used if you guess that a man is SLIGHTLY SHORT, that is, a little shorter than average. The next space on the scale would indicate that you thought the man was QUITE SHORT and finally the space closest to the word short should be used if you guess a man is VERY SHORT.

Example C shows the tall-short scale with a guess already indicated. When I heard man number 1 speak I guessed that he was just slightly taller than average so I placed my X in the space next to the center space on the tallside of the scale. Example D shows my guess about the height of man 2. Something about the way he sounded made me guess that he was QUITE SHORT so I put my X in the second space from the word short.

Now let's try a few for practice. Before you hear the voice, you will be told what number it is. Be sure to put your X for a particular voice on the scale next to the voice number [Voices 3, 4 and 5 were played and time was allowed for rating.] Are there any questions? (Pause)

OK let's try some more. On each of the next three

pages you will make guesses about 12 voices. On each page you will mark your guesses on a different scale. Write down your first impression and be sure that you put down an X each time you hear a voice. However, never put more than 1 X on the same scale. Be sure to put your X's in the middle of one of the spaces; never put an X on the dots.

Please turn the page and notice which adjective scale you willbe using to make guesses about the next 12 voices.

The above phrase followed the first two sets of twelve voiceson each type.

All voices were preceded by an announcement of the voice number and followed by approximately 10 seconds of silence.

APPENDIX B

Sample Instruction Sheet

NAME _____ SEX: M F

BIRTHDATE:
Month Day Year GRADE: _____

_____ _____ _____

Example A:

TALL ____:____:____:____:____:____:____ SHORT

Example B:

	VERY TALL	QUITE TALL	SLIGHTLY TALL	AVERAGE	SLIGHTLY SHORT	QUITE SHORT	VERY SHORT	
TALL	____	____	____	____	____	____	____	SHORT

Example C:

MAN
Number

1	TALL	____	____	X	____	____	____	SHORT

Example D:

2	TALL	____	____	____	____	____	X	SHORT

PRACTICE VOICES

3	TALL	____	____	____	____	____	____	SHORT
4	TALL	____	____	____	____	____	____	SHORT
5	TALL	____	____	____	____	____	____	SHORT

DO NOT TURN PAGE UNTIL TOLD TO DO SO

Sample Rating Sheet

MAN NUMBER		VERY UNFAIR	QUITE UNFAIR	SLIGHTLY UNFAIR	AVERAGE	SLIGHTLY FAIR	QUITE FAIR	VERY FAIR	
1	UNFAIR								FAIR
2	UNFAIR								FAIR
3	UNFAIR								FAIR

		VERY UNFAIR	QUITE UNFAIR	SLIGHTLY UNFAIR	AVERAGE	SLIGHTLY FAIR	QUITE FAIR	VERY FAIR	
4	UNFAIR								FAIR
5	UNFAIR								FAIR
6	UNFAIR								FAIR

		VERY UNFAIR	QUITE UNFAIR	SLIGHTLY UNFAIR	AVERAGE	SLIGHTLY FAIR	QUITE FAIR	VERY FAIR	
7	UNFAIR								FAIR
8	UNFAIR								FAIR
9	UNFAIR								FAIR

DO NOT TURN PAGE UNTIL TOLD TO DO SO

LIST OF REFERENCES

Allport, G. and Cantril, H. Judging personality from the voice. Journal of Social Psychology, 1934, 5, 37-55.

Barnard, J. W. The effects of anxiety on connotative meaning. Child Development, 1966, 37, 461-472.

Brandwin, M. A. Levels of reference in connotative meaning: A developmental study. Papers of the Michigan Academy of Science, Arts and Letters, 1965, 50, 343-353.

Cantril, H. and Allport, G. The Psychology of Radio. New York: Harper, 1935.

Costanzo, F. S., Markel, N. N. and Costanzo, P. R. Voice quality profile and perceived emotion. Journal of Counseling Psychology, 1969, 16, 267-270.

Dimitrovsky, L. The ability to identify the emotional meaning of vocal expressions at successive age levels. In Joel R. Davitz (Ed.) The Communication of Emotional Meaning. New York: McGraw-Hill, 1964.

Dittman, A. T. and Wynne, L. C. Linguistic techniques and the analysis of emotionality in interviews. Journal of Abnormal and Social Psychology, 1961, 63, 201-204.

Di Vesta, F. J. A normative study of 220 concepts rated on the semantic differential in grades 2 through 7. Journal of Genetic Psychology, 1966, 109, 205-229.

Di Vesta, F. J. and Dick, W. The test-retest reliability of children's ratings on the semantic differential. Educational and Psychological Measurement, 1966, 26, 605-610.

Donahoe, J. W. Changes in meaning as a function of age. Journal of Genetic Psychology, 1961, 99, 23-28.

Gates, G. S. The role of the auditory element in the interpretation of emotion. Psychological Bulletin, 1927, 24, 175.

Goodenough, E. W. Interest in persons as an aspect of sex differences in the early years. Genetic Psychology Monographs, 1957, 55, 287-323.

Kramer, E. Judgment of personal characteristics and emotions from nonverbal properties of speech. Psychological Bulletin, 1963, 60, 408-420.

Kramer, E. Personality stereotypes in voice: A reconsideration of the data. Journal of Social Psychology, 1964, 62, 247-251.

Lilly, R. S. A developmental study of the semantic differential. Doctoral Dissertation, Princeton University, 1965.

Mahl, G. F. and Schulze, G. Psychological research in the extralinguistic area. In T. A. Sebeok, A. S. Hayes and M. C. Bateson, (Eds.), Approaches to Semiotics. The Hague: Mouton and Company, 1964, pp. 51-124.

Maltz, H. E. Ontogenetic changes in the meaning of concepts as measured by the semantic differential. Child Development, 1963, 34, 667-674.

Markel, N. N. and Hamp, E. P. Connotative meanings of certain phoneme sequences. Studies in Linguistics, 1960, 15, 47-61.

Markel, N. N. The reliability of coding paralanguage: Pitch, loudness and tempo. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 306-308.

Markel, N. N. The relationship between voice quality profiles and MMPI profiles in psychiatric patients. Journal of Abnormal Psychology, 1969a, 74, 61-66.

Markel, N. N. Personal communication, 1969b.

Markel, N. N., Meisels, M. and Houck, J. E. Judging personality from voice quality. Journal of Abnormal and Social Psychology, 1964, 69, 458-463.

Markel, N. N. and Roblin, G. L. The effect of content and sex-of-judge on judgments of personality from voice. International Journal of Social Psychiatry, 1965, 11, 295-300.

Osgood, C. E., Suci, G. J. and Tannenbaum, P. H. The Measurement of Meaning. Urbana: University of Illinois Press, 1957.

Pear, T. H. Voice and Personality. London: Chapman and Hall, 1931.

Sapir, E. Speech as a personality trait. American Journal of Sociology, 1927, 32, 892-905.

Small, E. R. Age and sex differences in the semantic structure of children. Doctoral Dissertation, University of Michigan, 1958.

Stricker, and Zax. Intelligence and semantic differential discriminability. Psychological Reports, 1966, 18, 775-778.

Thorndike, E. L. and Lorge, I. The Teacher's Word Book of 30,000 Words. New York: Bureau of Publications, Teacher's College, Columbia University, 1944.

Trager, G. L. Paralanguage: A first approximation. Studies in Linguistics, 1958, 13, 1-12.

Vargas, R. A study of certain personality characteristics of male college students who report frequent positive experiencing and behaving. Doctoral Dissertation, University of Florida, 1968.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill Book Company, 1962.

BIOGRAPHICAL SKETCH

Judith Ann Phillis was born in Uniontown, Pennsylvania, on March 22, 1942. She was graduated from Susquehanna Township High School, Harrisburg, Pennsylvania, in June of 1960. In the fall of that year she entered Carnegie Institute of Technology (Carnegie-Mellon University), Pittsburgh, Pennsylvania, from which she was graduated with the degree of Bachelor of Science in June, 1964.

From June, 1964, through August, 1965, Miss Phillis was employed as a research assistant at the Craig House for Children in Pittsburgh, Pennsylvania. In September, 1965, she began graduate study in psychology at the University of Florida. From June, 1967, until August, 1967, she was employed as a laboratory technician at the Communication Sciences Laboratory, University of Florida. She received the degree Master of Arts in August, 1967, and since that time she has pursued her work toward the degree of Doctor of Philosophy.

This dissertation was prepared under the direction of the chairman of the candidate's supervisory committee and has been approved by all members of that committee. It was submitted to the Dean of the College of Arts and Sciences and to the Graduate Council, and was approved as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

December, 1969

E. Raffin Jones
Dean, College of Arts and Sciences

Dean, Graduate School

Supervisory Committee:

Chairman

W. S. Schumacher

Norma D. Markel

Jerry VanDeKieft

H. Brueppacher

Madeline Carey